Journal Name

ARTICLE TYPE

Cite this: DOI: 10.1039/xxxxxxxxxx

Accurate empirical rovibrational energies and transitions of $H_2{}^{16}O$

Roland Tóbiás,[a] Tibor Furtenbacher,[a] Jonathan Tennyson,[b] and Attila G. Császár*[,c]

Received Date
Accepted Date

DOI: 10.1039/xxxxxxxxxx

www.rsc.org/journalname

Several significant improvements are proposed to the MARVEL (Measured Active Rotation-al-Vibrational Energy Levels) protocol. The most important algorithmic changes include the use of groups of transitions, blocked by their estimated experimental (source segment) uncertainties, an inversion and weighted least-squares refinement procedure based on sequential addition of blocks of decreasing accuracy, the introduction of spectroscopic cycles during the refinement process, automated recalibration, synchronization of the ground-state combination difference relations to reduce residual uncertainties in the resulting dataset of empirical (MARVEL) energy levels, and improved classification of the lines and energy levels based on their accuracy and dependability. The resulting protocol, through handling a large number of measurements of similar accuracy, retains, or even improves upon, the best reported uncertainties of the spectroscopic transitions employed. To show its advantages, the extended MARVEL protocol is applied for the analysis of the complete set of highly accurate $H_2{}^{16}O$ transition measurements. As a result, almost 300 highly accurate energy levels of $H_2{}^{16}O$ are reported in the energy range of 0–6000 cm^{-1}. Out of the 15 vibrational bands involved in accurately measured rovibrational transitions, the following three have definitely highly accurate empirical rovibrational energies of 8–10 digits of accuracy: $(v_1\ v_2\ v_3) = (0\ 0\ 0)$, (0 1 0), and (0 2 0), where v_1, v_2, and v_3 stand for the symmetric stretch, bend, and antisymmetric stretch vibrational quantum numbers. The dataset of experimental rovibrational transitions and empirical rovibrational energy levels assembled during this study, both with improved uncertainties, is considerably larger and considerably more accurate than the best previous datasets.

1 Introduction

In the era of optical frequency combs and frequency-comb spectroscopy[1,2] it has become increasingly realistic to determine experimental line positions with an accuracy on the order of 10 kHz or even better. Previously, such experimental accuracy was reserved to pure rotational transitions measured in the microwave (MW) region of the electromagnetic spectrum but nowadays there is growing evidence that this accuracy can be extended all the way from the MW to the near infrared (NIR) region[3–5] and beyond.[6,7]

A related relevant development in theoretical high-resolution molecular spectroscopy has been the introduction of spectroscopic networks,[8–10] and the Measured Active Rotational-Vibrational Energy Levels (MARVEL) scheme.[11,12] The MARVEL algorithm and code, originally presented in Ref. 11 and 12, allows the efficient determination of accurate empirical energy levels with well-defined uncertainties from a set of measured and assigned rovibronic transitions, the global analysis of measured spectra, and the validation of related spectroscopic datasets and databases.[13] MARVEL has been employed for a number of molecular systems, including $^{12}C_2$,[14] nine isotopologues of water,[15–19] three isotopologues of SO_2,[20] three isotopologues of H_3^+,[21,22] $^{14}NH_3$,[23] and parent ketene.[24] A web-based version of the MARVEL code has allowed the involvement of high-school students[25] in spectroscopic research projects leading to published studies on $^{48}Ti^{16}O$,[26] $^{90}Zr^{16}O$,[27] $H_2{}^{32}S$,[28] and $^{12}C_2H_2$,[29] molecules of considerable astronomical interest.

Energy levels determined using the MARVEL procedure are increasingly being used to improve not only transition wavenumbers[30] but also partition functions[31] present in standard databases. Just considering water, MARVEL energy levels have been used extensively by the ExoMol project[32] to improve predicted transition frequencies in line lists for $H_2{}^nO$ (n =

[a] Laboratory of Molecular Structure and Dynamics, Institute of Chemistry, ELTE Eötvös Loránd University and MTA-ELTE Complex Chemical Systems Research Group, P.O. Box 32, H-1518 Budapest 112, Hungary
[b] Department of Physics and Astronomy, University College London, London WC1E 6BT, United Kingdom
[c] MTA-ELTE Complex Chemical Systems Research Group, P.O. Box 32, H-1518 Budapest 112, Hungary; E-mail: csaszar@caesar.elte.hu

16, 17, 18).[33,34] Similarly, the most recent edition of the HITRAN database, HITRAN2016,[30] makes use of MARVEL energy levels for $H_2{}^nO$ ($n = 16, 17, 18$),[35,36] as well as for the deuterated isotopologues.[37] MARVEL energy levels formed a key part of a procedure used to greatly improve predicted energy levels for water isotopologues.[33] Empirical (MARVEL) rovibrational energies have also been used to generate the most accurate partition functions available for $H_2{}^{16}O$[38] and heavy water (and its three constituent isotopologues).[39] As they are present in a large number of spectra, water lines can also be used for calibration in different spectral regions.[40,41]

The second-generation MARVEL code[12] runs extremely fast, treats transition data, and the related matrices (*vide infra*) on the order of 100 000 in less than a minute on a single CPU, and it is platform independent. This efficiency of the MARVEL code facilitates enhancements, such as those discussed below, of the MARVEL algorithm.

In all the MARVEL studies cited above accurate empirical rovibronic energy levels have been determined from the simultaneous treatment of all measured transitions. Nevertheless, it has repeatedly been observed that when transitions based on the empirical (MARVEL) energy levels and their uncertainties were compared to the best measurements slight distortions and occasionally unnecessarily large, but sometimes too small, uncertainties characterize the MARVEL data. These inconsistencies, at least in part, are the result of (a) the inclusion of many transitions of orders of magnitude lower accuracy than the best measurements in the MARVEL inversion and refinement procedure, and (b) the large number of inaccurately measured transitions connecting more accurately measured transitions to the lowest energy levels of the principal components of the experimental SN. In this study we attempt to devise a MARVEL-based protocol which retains the accuracy of the best measurements while still working with as complete spectroscopic networks as feasible during the inversion and refinement process.

Next, let us review the state-of-the-art of high-resolution spectroscopy for the $H_2{}^{16}O$ isotopologue of water, the subject molecule of our feasibility study in which the extended MARVEL protocol is employed to gain highly accurate rovibrational energy levels meeting or even exceeding the accuracy of the best experiments.

The highest-quality database of highly accurate rovibrational energy levels and transitions of water vapor is the database maintained at the Jet Propulsion Laboratory (JPL).[42,43] Nowadays the $H_2{}^{16}O$ dataset of JPL is based on a study of Lanquetin *et al.*[44] In particular, the JPL energy levels of $H_2{}^{16}O$ are exactly those reported in Ref. 44. The JPL transitions are the results of an effective Hamiltonian fit of a considerable number of transitions, as detailed on the JPL website.[43] The most comprehensive evaluation of measured water transitions was performed by an IUPAC Task Group.[15–19] Ref. 19 gives a summary of this work. Part III of this series[17] contains a validated and recommended set of measured $H_2{}^{16}O$ transitions (about 200 000) and empirical energy levels (about 20 000), based on experimental data available prior to 2013. These sets were used to update the water data in HITRAN2016.[30] Since 2013, there have been many new studies of water spectra and, in particular, the use of optical-frequency-comb-based measurements to determine very accurate frequencies for selected transitions in the infrared, see, *e.g.*, Refs. 5, 45, 46. The advent and availability of new, precise experimental techniques act as an impetus to further improve the MARVEL treatment of measured transitions allowing the determination of highly accurate empirical energy levels.

As of today no *ortho*-*para* rovibrational transitions have been observed in water vapor.[66] This has the consequence that water spectra are related to two principal components[10] of the spectroscopic network of $H_2{}^{16}O$, those of *ortho*- and *para*-$H_2{}^{16}O$. Note that the number $q = v_3 + K_a + K_c$ is even for *para* and odd for *ortho* rovibrational states, where v_3 is the vibrational quantum number corresponding to the antisymmetric stretch motion, while K_a and K_c are the usual rigid-rotor quantum numbers of an asymmetric-top molecule.

As emphasized in Ref. 41, despite the fact that *ortho* lines may have three times higher intensity than *para* lines (and thus about two times more of them have been determined experimentally[17]), the best frequency standards correspond to *para*-$H_2{}^{16}O$, as there is no hyperfine splitting of the *para*-$H_2{}^{16}O$ lines and the

Table 1 The 13 astrophysically most important water ($H_2{}^{16}O$) lines recommended by the International Astronomical Union and their different experimental determinations. The experimental (Expt.) frequencies of multiple measurements follow the order of their increased uncertainties. In the assignment column, the standard spectroscopic notation $J'_{K'_a,K'_c} \leftarrow J''_{K''_a,K''_c}$ is used, where $J'_{K'_a,K'_c}$ and $J''_{K''_a,K''_c}$ are the rotational labels for the upper and lower energy levels, respectively. All transitions belong to the ground vibrational state.

Approx. Freq./GHz	Component	Assignment	extMARVEL/kHz	01LaCoCa[44]/kHz	Expt./kHz
22.235	*para*	$6_{16} \leftarrow 5_{23}$	22235079.85(6)	22235007(4240)	22235079.85(6)[47]
					22235080(15)[48]
					22235200(132)[49]
					22235220(154)[50]
					22235000(190)[51]
183.310	*ortho*	$3_{13} \leftarrow 2_{20}$	183310090.4(1)	183309897(2544)	183310090.6(1)[52]
					183310087(4)[53]
					183310117(29)[54]
					183310150(65)[48]
					183310200(190)[51]
					183311300(1330)[55]
325.153	*para*	$5_{15} \leftarrow 4_{22}$	325152899(2)	325153101(3392)	325152899(2)[53]
					325152888(13)[56]
					325152919(22)[54]
					325153700(882)[51]
380.197	*para*	$4_{14} \leftarrow 3_{21}$	380197359.8(6)	380197395(3610)	380197359.8(6)[57]
					380197356(4)[53]
					380197365(13)[56]
					380197372(17)[54]
					380196800(621)[51]
439.151	*para*	$6_{43} \leftarrow 5_{50}$	439150794.8(6)	439151582(5528)	439150794.8(6)[57]
					439150795(2)[53]
					439150812(19)[54]
448.001	*para*	$4_{23} \leftarrow 3_{30}$	448001077.5(6)	448001155(3610)	448001077.5(6)[57]
					448001075(17)[54]
					448000300(904)[51]
474.689	*ortho*	$5_{33} \leftarrow 4_{40}$	474689108(2)	474689879(3816)	474689108(2)[53]
					474689127(21)[54]
556.936	*para*	$1_{10} \leftarrow 1_{01}$	556935987.6(6)	556935841(1824)	556935987.7(6)[57]
					556935985(3)[58]
					556935995(8)[53]
					556936002(17)[54]
					556935819(186)[59]
					556935800(190)[51]
					556935800(206)[60]
620.700	*para*	$5_{32} \leftarrow 4_{41}$	620700954.9(6)	620701697(4664)	620700954.9(6)[57]
					620700950(34)[61]
					620700844.1(122)[62]
					620700807(163)[54]
752.033	*ortho*	$2_{11} \leftarrow 2_{02}$	752033113(15)	752032978(2544)	752033113(15)[63]
					752033104(45)[59]
					752033227(121)[54]
					752033300(190)[51]
916.172	*ortho*	$4_{22} \leftarrow 3_{31}$	916171581(61)	916171448(3392)	916171580(61)[64]
					916171582(63)[65]
					916171405(155)[59]
970.315	*ortho*	$5_{24} \leftarrow 4_{31}$	970315021(21)	970315165(3392)	970315022(21)[65]
					970315020(21)[64]
					970314968(76)[59]
987.927	*ortho*	$2_{02} \leftarrow 1_{11}$	987926762(21)	987926473(2341)	987926764(21)[65]
					987926760(21)[64]
					987926743(23)[59]

minimum-energy level of the *para* PC can properly be set to zero with zero uncertainty.

High-resolution radial-velocity-shift transition measurements used to detect molecular species in the atmospheres of exoplanets[67,68] has greatly increased the need for accurate laboratory transition frequencies. While this technique has been used successfully to identify water in exoplanets[69,70] with high confidence, standard water line lists used for exoplanet modelling appear to be not sufficiently accurate for this task.[71] It is also true that many of the most important line positions have been measured sufficiently accurately. In its extended list[72] of the astrophysically most important spectral lines, the International Astronomical Union listed 13 $H_2{}^{16}O$ lines, with assignments and rest frequencies recalled in Table 1. For all these frequencies several independent, highly accurate experimental determinations are available (see Table 1). Nevertheless, not all of these transitions are part of cycles satisfying the law of energy conservation[13] to the accuracy of the best measurements, which calls for OFC-based remeasurement of certain rovibrational transitions on the ground vibrational state of water.

The present study has been executed with two principal goals in mind. First, we wanted to search for improvements in the MARVEL protocol,[11,12] with particular emphasis on a better reproduction of the most accurately measured rovibrational transitions. Second, to test the feasibility of the improved protocol, we chose to study $H_2{}^{16}O$, the admittedly most important polyatomic molecule for high-resolution spectroscopy, where the accurate knowledge of the rovibrational energy levels and transitions is important in a number of scientific and engineering applications. Therefore, first a detailed description of the improved MARVEL protocol is given in Section 2. To aid the reader of the methodological section, some of the more technical details are moved to appendices associated with the subsections. Those interested only in the spectroscopic results of this study, and not the way they were determined, can skip Section 2. Then, Section 3 discusses the $H_2{}^{16}O$ experimental input data used, while Section 4 presents results of our improved MARVEL analysis of a large number of old and new sources.[45–65,73–102] This analysis also includes comparisons with previously determined accurate water levels and lines.[44,59,94] The paper ends with some concluding remarks in Section 5.

2 Methodology

2.1 Spectroscopic networks

Sets of rovibrational transitions, whether they are measured or computed, can be treated as building blocks of *spectroscopic networks* (SN, see Refs. 8–13), whereby (a) the *vertices* are the energy levels, (b) the *edges* correspond to the transitions, oriented from the lower energy level to the upper one (independently whether the transition was recorded in absorption, emission, or by means of an action spectroscopy), and (c) the (positive) *edge weights* are the wavenumbers of the transitions. Note that many other weighting schemes (*e.g.*, weighting by line intensities) can be utilized in practical applications of SNs.

If SNs are formed by experimental transitions, called *experimental SNs*, the edge weights should be associated with measurement uncertainties. However, in many data sources, instead of providing line-by-line uncertainties, only an average "expected" accuracy is provided, usually corresponding to the most intense unblended lines, *i.e.*, to a best-case measurement scenario. Based on these experimental line positions and approximate uncertainties —together with assignments to the lower and upper energy levels of the transition—, one is able to derive *empirical (MARVEL) energy levels* with uncertainties via the MARVEL (Measured Active Rotational-Vibrational Energy Levels) procedure,[11,12] utilizing a weighted least-squares technique (*e.g.*, robust reweighting[103]) and the *Rydberg–Ritz combination principle*.[104]

The detailed analysis of experimental SNs is very important as one would like to treat all the measured rovibrational lines of a molecule simultaneously and make the validated ones available to spectroscopists and spectroscopic database developers. For this purpose, it is necessary to scrutinize (a) the *components* (sets of energy levels not connected by any measured transition), (b) the *bridges* (transitions whose deletion increases the number of components), and (c) the *cycles* (collections of connected edges within which every vertex has two neighboring vertices) characterizing the given experimental SN.

Since the energy levels of the different components are not connected, during the MARVEL analysis we are forced to set the lowest-lying energy level (*core*) of each component to zero. If the core of a component corresponds to the lowest-energy level of a particular nuclear-spin isomer of the molecule examined, this component is a *principal component* (PC), otherwise it is called a *floating component* (FC). Clearly, one is most interested in the detailed characterization of those energy levels and transitions which are part of PCs.

Due to the fact that bridges of experimental SNs may compromise the accuracy of the empirical energy levels determined during a MARVEL analysis, they require special attention. That is, if a bridge is determined incorrectly or inaccurately, the energy levels connected by it to the core of their component will be shifted or scattered. Energy levels "behind" a bridge cannot be known more accurately than the bridge itself, a considerable hindrance in the derivation of highly accurate empirical energy levels.

Cycles are extremely useful when compatibility of the transitions and their associated uncertainties obtained in different groups under often widely differing experimental conditions are examined.[105] It is important to check, for all the cycles of the SN, whether the *law of energy conservation* (LEC, see Ref. 13) is satisfied within the experimental accuracy. Accordingly, in the cases where (a) a transition is assigned improperly, (b) its wavenumber is measured inaccurately, or (c) the uncertainty of this wavenumber is underestimated, the *discrepancy* (absolute signed sum of the line positions) becomes higher than the *threshold* (the sum of the uncertainties) in the given cycle, indicating a conflict among the related transitions.

All the cycles of an experimental SN can be expressed in a *cycle basis* (CB) with the symmetric differences of the *basic cycles* in the CB, which contains all non-bridge lines of this SN. Accordingly, construction of such CBs and evaluation of their entries could be sufficient to test the compatibility of the measured rovibrational

lines in light of the LEC. For details about the use of CBs to explore inconsistencies in SNs, see Ref. 13. Note also that in contrast to Ref. 13, cycles of length 2, which correspond to repeated measurements, are also permitted here, as this extension makes the present formalism simpler than the previous one.[13]

Subnetworks are of great importance during the treatment of SNs. These derived graph structures are (a) represented with a *participation matrix*, $\mathbf{P} = \mathrm{diag}(P_1, P_2, ..., P_{N_T})$, where N_T is the *number of lines* in the SN, and $P_i = 1$ if the ith transition of the SN is inserted in the subnetwork, otherwise $P_i = 0$, and (b) filled with *all* the energy levels of the SN, among which there may also be *isolated nodes* (vertices without inserted lines).

Due to the presence of some outliers, which should be excluded from the database of transitions during our analyses (see also section 2.3.2), it is necessary to introduce the *leading subnetwork* of the SN, denoted with $\mathscr{N}_{LS}$, which contains all the non-excluded lines of the database. In the present description, $\mathbf{P}$ will always denote the participation matrix assigned to the $\mathscr{N}_{LS}$ subnetwork.

2.2 The MARVEL procedure

To understand the improvements proposed in this study better, first the traditional MARVEL approach,[11,12] upon which our novel blockMARVEL algorithm (*vide infra*) is built, is recalled. During a MARVEL analysis of measured rovibronic transitions the following objective function is minimized:

$$\Omega(\mathbf{E}) = (\sigma - \mathbf{RE})^T \mathbf{PW} (\sigma - \mathbf{RE}), \quad (1)$$

where (a) $\mathbf{P}$ is the participation matrix of $\mathscr{N}_{LS}$, (b) $\mathbf{E} = \{E_1, E_2, ..., E_{N_L}\}^T$ is the vector of energies of the N_L energy levels, (c) $\sigma = \{\sigma_1, \sigma_2, ..., \sigma_{N_T}\}^T$ is the vector of wavenumbers of the N_T transitions, (d) $\mathbf{R}$ is the *Ritz matrix* (see also Eq. (3) of Ref. 11), (e) $\mathbf{W} = \mathrm{diag}(w_1, w_2, ..., w_{N_T})$ is a diagonal *weight matrix* with $w_i = \delta_i^{\eta}$, (f) δ_i is the uncertainty of σ_i, and (g) η is a nonpositive exponent, set to -2 in our calculations.

Since the minimum of $\Omega(\mathbf{E})$, denoted by $\bar{\mathbf{E}} = \{\bar{E}_1, \bar{E}_2, ..., \bar{E}_{N_L}\}^T$, is not unique,

$$\bar{E}_{\mathrm{core}(1)} = \bar{E}_{\mathrm{core}(2)} = ... = \bar{E}_{\mathrm{core}(N_c)} = 0 \quad (2)$$

should be set for the cores indexed with core(1), core(2), ..., core(N_c), where N_c is the number of components in $\mathscr{N}_{LS}$. Furthermore, $\bar{\mathbf{E}}$ must also satisfy the relation

$$\mathbf{G}\bar{\mathbf{E}} = \mathbf{F}, \quad (3)$$

where $\mathbf{G} = \mathbf{R}^T\mathbf{PWR}$ is the *weighted Gram–Schmidt matrix* of $\mathbf{R}$, and $\mathbf{F} = \mathbf{R}^T\mathbf{PW}\sigma$ is the *vector of free terms*. Note that Eqs. (2)–(3) can be solved with *Cholesky decomposition*, using, for example, the standard Eigen package,[106] and can be reduced by transforming $\mathbf{G}$ into a block diagonal form, which consists of N_c independent diagonal blocks. Note that during the testing of the code we experimented with several decomposition procedures, but no significant differences were found. Nevertheless, there is a factor which influences substantially the stability of the solution, commented upon in Section 2.2 (right after Eq. (5)).

In practice, the indices of the cores of $\mathscr{N}_{LS}$ are not necessarily known to the user. Then, from any solution, $\mathbf{E}' = \{E'_1, E'_2, ..., E'_{N_L}\}^T$, of Eq. (3), the core($i$) index can be obtained as

$$\mathrm{core}(i) = \underset{\substack{j=1 \\ \mathrm{comp}(j)=i}}{\overset{N_L}{\mathrm{argmin}}} E'_j \quad (4)$$

for all $1 \le i \le N_c$, where comp(j) is the *component index* of the jth energy level, *i.e.*, the index of the component containing this energy level. Based on the core indices, the $\bar{E}_j$ values can be expressed for all $1 \le j \le N_L$ as

$$\bar{E}_j = E'_j - E'_{\mathrm{core}(j)}. \quad (5)$$

In what follows, we always assume that the core indices of $\mathscr{N}_{LS}$ are available.

To eliminate the linear dependencies from the block-diagonal form of $\mathbf{G}$, one has to leave out a (row,column) pair from each block. Although, in principle, these pairs could be chosen arbitrarily, in practice it is best to select those (row,column) combination which are associated with the largest diagonal entry of the corresponding blocks, in order to improve the numerical stability of the solution.

One must note in passing that measurement uncertainty is a combination of precision and absolute accuracy. Precision of measured line positions varies considerably according to a number of factors: (a) type and quality of the spectrometer, (b) the spectral resolution relative to the observed line width, (c) the signal to noise ratio, (d) the choice and especially the control of experimental conditions, (e) the complexity of the spectrum, and (f) the line retrieval methods employed. Since all of these factors must be considered as carefully as possible, it is not at all surprising that experimentalists measuring a large number of lines often refrain from reporting dependable line-by-line uncertainties.

As soon as $\bar{\mathbf{E}}$ has been determined, it is necessary to examine the *(fitting) residuals* ($\Delta_i = \sigma_i - \bar{E}_{\mathrm{up}(i)} + \bar{E}_{\mathrm{low}(i)}$) and the *(fitting) defects* ($d_i = |\Delta_i| - \delta_i$), where up($i$) and low($i$) are the indices of the upper and lower energy levels of the ith transition, respectively. $\mathscr{N}_{LS}$ is called *consistent* if the largest defect of $\mathscr{N}_{LS}$, $d_{\max}$, is not positive. In the case of $d_{\max} > 0$, we have to either exclude certain lines or increase their uncertainties in order to ensure the consistency of $\mathscr{N}_{LS}$.

2.3 The extended MARVEL (extMARVEL) approach

The approach outlined here is an extension of the MARVEL procedure,[11,12] which proved to be a powerful tool to obtain correct and reliable empirical rovibrational energy levels and associated uncertainties in a number of applications, as detailed in the Introduction. Our improved protocol is depicted in Fig. 1, the following subsections explain each of the schemes shown in Fig. 1.

2.3.1 Isolation of source segments (ISS)

Having collated the transitions into a SN, good estimates for the uncertainties of the line positions are required. Fortunately, in practice it is often sufficient to find a suitable uncertainty for each subset of transitions of the same (data) source which can be modeled using identical uncertainties, at least as a first approximation.

Then, one needs to divide the transitions of the sources into *segments*, in which the uncertainties of the wavenumbers are considered to be (approximately) within a factor of 10. A particular

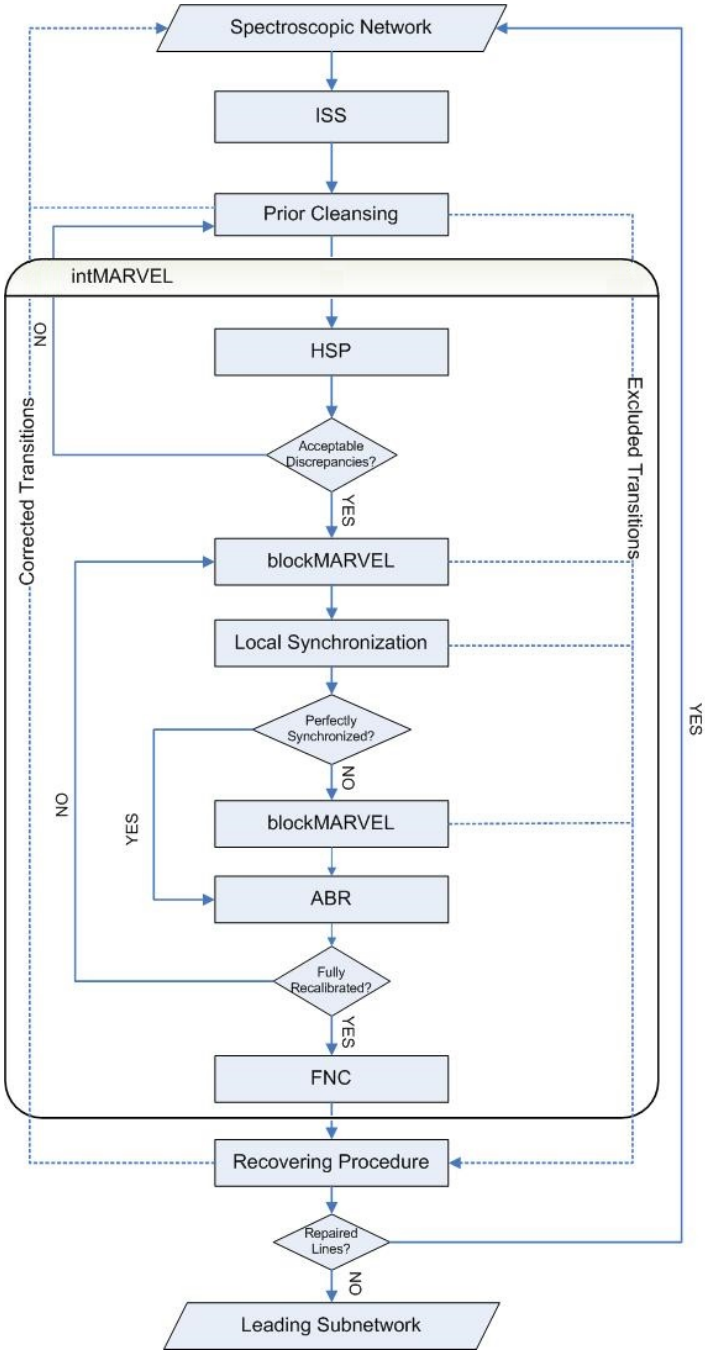


Fig. 1 The extended MARVEL (extMARVEL) approach, see text for an explanation of each step.

segment is denoted in this study with (a) the tag of its source if this source contains a single segment, or (b) a character sequence obtained by the concatenation of the source tag with one of the strings '_S2', '_S3', etc. By convention, (a) segments are indexed in increasing order of their uncertainty, and (b) the first segment, '_S1', is never written out explicitly.

It is important to establish the *estimated segment uncertainties* (ESU) with extreme caution to avoid the problem of under- and overutilization of the data. If explicit information on the accuracy is not available from the sources, we use the simple approximate relations 'uncertainty = resolution / 10' and 'uncertainty = resolution' for the unblended and blended lines, respectively. (It would help spectroscopic database developers if experimentalists published their lines partitioned at least into blended and unblended transitions and reported average uncertainty estimates for at least these two categories.)

At the end of this process, two input files need to be created for further analyses. The first one is the transition database in MARVEL format[11,12] with the distinction that the "uncertainty" column is missing, or if present, it will not be evaluated. The second one is the list of the source segments with their ESUs and a ternary flag for each segment. The value of this flag is (a) 0, if the given segment must not be subjected to recalibration using a single recalibration factor, (b) −1, if this segment should be recalibrated with the same factor as the segment in the previous record of the segment file, or (c) 1, otherwise.

2.3.2 Prior cleansing

Before the empirical energy levels are determined it is mandatory to perform an extensive cleansing of the collated dataset to weed out clear outliers. As to inconsistencies concerning the dataset, both *errors* (*misprints* or *transcription errors*) and *inaccuracies* may occur (a) in the wavenumbers, (b) in the line assignments, and (c) in the ESU values of the segments. All these problems need to be identified. For this purpose, several checks must be carried out, including (a) a test whether the selection rules hold properly for the input transitions, (b) use of the ECART (Energy Conservation Analysis of Rovibronic Transitions, see Ref. 13) algorithm to list and delete the incorrect cycles, or (c) an analysis of the fitting residuals (see Sec. 2.2) derived from the MARVEL procedure.

2.3.3 Integrated MARVEL (intMARVEL) analysis

The integrated MARVEL (intMARVEL) algorithm (see Fig. 1) is a fully automated procedure able to (a) generate refined segment uncertainties (RSU) from the ESU values, (b) determine the energy levels of the SN based on their most accurate lines, (c) exclude outlier transitions from the SN, (d) calculate the individual wavenumber uncertainties of the lines originating from the RSU values, (e) synchronize the *ground-state combination difference* (gsCD) relations, (f) recalibrate ill-calibrated segments, and (g) assess the lines, segments, and energy levels on the basis of their accuracy and dependability.

This procedure was programmed into a code, written in the C++ language, called `intMARVEL`, which requires the two input files mentioned in Sec. 2.3.1. The output of the `intMARVEL` code contains the transitions, the energy levels, and the segments characterized, all in separate text files. A detailed description of the theoretical background of the intMARVEL algorithm is given in Sec. 2.4 and the corresponding appendices.

2.3.4 Recovering procedure

Once the intMARVEL analysis has been completed, the transitions excluded during certain stages of the extMARVEL algorithm need to be re-assessed and, if possible, corrected. In order to achieve the data correction, the wavenumbers and assignments of the excluded lines are to be checked carefully based on the original

sources.

After this revision, reassignments can be made for the non-corrected, excluded lines, using (a) the wavenumber-sorted experimental transition dataset, and (b) a MARVEL or an effective Hamiltonian (EH) linelist ignoring weak and forbidden transitions. At the end of the recovering step, the extMARVEL procedure should be repeated with the corrected lines until no transitions can be "repaired".

2.4 Background of the intMARVEL analysis

2.4.1 Hieararchical segment perturbation (HSP)

To obtain the RSU values, the *hierarchic segment perturbation* (HSP) scheme is introduced. During HSP, for all the s segments, (a) a $\mathscr{N}_{\rm per}$ *perturbation subnetwork* is constructed by including all the non-excluded lines from s and those segments which are much more accurate than s (*hierarchical perturbation*), and (b) the RSU value is provided for s from the discrepancies and thresholds of those basic cycles in a given CB of $\mathscr{N}_{\rm per}$ which contain transitions from s (*representative cycles*). If there are too few representative cycles for the refinement of a particular ESU, then the corresponding RSU will be identical to this ESU value. The consecutive stages of the HSP procedure are described in Appendix A.

2.4.2 BlockMARVEL refinement

Since inaccurately measured (or inaccurately represented) wavenumber entries may deteriorate the accuracy of the estimated energy levels, a novel procedure, called the *constrained MARVEL scheme*, was designed with which these harmful effects can be avoided or at least substantially reduced. This technique ensures that the residuals derived from a given $\mathscr{N}_{\rm p}$ (previous) subnetwork of $\mathscr{N}_{\rm LS}$, *via* the MARVEL algorithm, should remain unaffected when the MARVEL analysis is repeated for a $\mathscr{N}_{\rm a}$ (actual) subnetwork of $\mathscr{N}_{\rm LS}$, which also contains $\mathscr{N}_{\rm p}$. By definition, the cores of $\mathscr{N}_{\rm a}$ are the cores of $\mathscr{N}_{\rm p}$, as well.

Designate the indices of the cores in $\mathscr{N}_{\rm p}$ and $\mathscr{N}_{\rm a}$ with $\mathrm{core_p}(1)$, $\mathrm{core_p}(2)$, ..., $\mathrm{core_p}(N_{\rm p,c})$ and $\mathrm{core_a}(1)$, $\mathrm{core_a}(2)$,..., $\mathrm{core_a}(N_{\rm a,c})$, respectively, where $N_{\rm p,c}$ and $N_{\rm a,c}$ are the number of components in $\mathscr{N}_{\rm p}$ and $\mathscr{N}_{\rm a}$, respectively. Let $\mathbf{P}^{\langle\mathscr{N}_{\rm a}\rangle}$ be the participation matrix of $\mathscr{N}_{\rm a}$. With the notation introduced, Eqs. (2)–(3) can be reformulated for $\mathscr{N}_{\rm a}$ as follows:

$$\bar{E}_{\mathrm{core_a}(1)} = \bar{E}_{\mathrm{core_a}(2)} = \ldots = \bar{E}_{\mathrm{core_a}(N_{\rm a,c})} = 0 \tag{6}$$

$$\mathbf{G}^{\langle\mathscr{N}_{\rm a}\rangle}\bar{\mathbf{E}} = \mathbf{F}^{\langle\mathscr{N}_{\rm a}\rangle}, \tag{7}$$

where $\mathbf{G}^{\langle\mathscr{N}_{\rm a}\rangle} = \mathbf{R}^{\rm T}\mathbf{P}^{\langle\mathscr{N}_{\rm a}\rangle}\mathbf{W}\mathbf{R}$ and $\mathbf{F}^{\langle\mathscr{N}_{\rm a}\rangle} = \mathbf{R}^{\rm T}\mathbf{P}^{\langle\mathscr{N}_{\rm a}\rangle}\mathbf{W}\sigma$. Since the residuals produced from $\mathscr{N}_{\rm p}$ with the MARVEL technique should remain unchanged, the following constraints must be enforced for all $1 \le i \le N_{\rm L}$:

$$\bar{E}_i - \bar{E}_{\mathrm{core_p}(\mathrm{comp_p}(i))} = \beta_i, \tag{8}$$

where $\mathrm{comp_p}(i)$ is the component index of the ith energy level in $\mathscr{N}_{\rm p}$ and β_i is the *bound* for $\bar{E}_i$. (Bounds correspond to empirical energy levels derived from $\mathscr{N}_{\rm p}$ using the MARVEL analysis.) In other words, only the cores of $\mathscr{N}_{\rm p}$ are linearly independent variables, all the other energy levels can be expressed from Eq. (8). Accordingly, these constraints can also be written in a matrix-vector form:

$$\bar{\mathbf{E}} = \mathbf{C}\bar{\mathbf{H}} + \beta, \tag{9}$$

where $\bar{\mathbf{H}} = \{\bar{H}_1, \bar{H}_2, \ldots, \bar{H}_{N_{\rm p,c}}\}^{\rm T}$ with $\bar{H}_j = \bar{E}_{\mathrm{core_p}(j)}$ and $\mathbf{C} = \{c_{ij}\}$ with

$$c_{ij} = \begin{cases} 1, \text{ if } j = \mathrm{comp_p}(i), \\ 0, \text{ otherwise.} \end{cases} \tag{10}$$

Applying Eqs. (7)–(9) and after some algebraic manipulations, the so-called *constrained MARVEL equations*,

$$\bar{H}_1 = \bar{H}_2 = \ldots = \bar{H}_{N_{\rm a,c}} = 0, \tag{11}$$

$$\mathbf{G}_{\mathbf{C}}^{\langle\mathscr{N}_{\rm a}\rangle}\bar{\mathbf{H}} = \mathbf{F}_{\mathbf{C}}^{\langle\mathscr{N}_{\rm a}\rangle}, \tag{12}$$

can be deduced, where $\mathbf{G}_{\mathbf{C}}^{\langle\mathscr{N}_{\rm a}\rangle} = \mathbf{C}^{\rm T}\mathbf{G}^{\langle\mathscr{N}_{\rm a}\rangle}\mathbf{C}$ and $\mathbf{F}_{\mathbf{C}}^{\langle\mathscr{N}_{\rm a}\rangle} = \mathbf{C}^{\rm T}(\mathbf{F}^{\langle\mathscr{N}_{\rm a}\rangle} - \mathbf{G}^{\langle\mathscr{N}_{\rm a}\rangle}\beta)$.

Note that $\mathbf{G}_{\mathbf{C}}^{\langle\mathscr{N}_{\rm a}\rangle}$ is a much smaller matrix than $\mathbf{G}$; thus, solving Eqs. (11)–(12) is much less expensive than solving Eqs. (2)–(3). In fact, for a database of approximately 260 000 $H_2{}^{16}O$ transitions, the traditional MARVEL and the constrained MARVEL were executed with fixed uncertainties. For the latter procedure, transitions were divided into eight blocks, indexed with -9, -8, -7, -6, -5, -4, -3, and -2, which contained 1, 9, 45, 666, 8304, 71990, 174965, and 3626 lines, respectively. The computational time concerning the older and the newer approaches was 4.5 and 17.8 s on a single processor, respectively, which corresponds to a speed-up of about four. Nevertheless, in contrast to the $\mathbf{G}$ matrix, $\mathbf{G}_{\mathbf{C}}^{\langle\mathscr{N}_{\rm a}\rangle}$ is not diagonally dominant; thus, a pivoting strategy must be employed to yield the solution of Eqs. (11)–(12). Once $\bar{\mathbf{H}}$ is determined, $\bar{\mathbf{E}}$ can easily be calculated by means of Eq. (9).

Utilizing Eqs. (11)–(12), a *divide-and-conquer*-style method (*blockMARVEL procedure*) was designed, whose successive steps are listed in Appendix B. During this refinement, (a) lines are divided into blocks according to the orders of magnitude of the associated RSUs, (b) $\mathscr{N}_{\rm a}$ is extended *block by block* through linking the transitions from the upcoming block to this subnetwork, and (c) the uncertainties of the added lines are adjusted to achieve consistency within $\mathscr{N}_{\rm a}$. At the end of this process, accurate *constrained empirical energy levels* and reliable wavenumber uncertainties are obtained, provided that (a) the gsCD relations are synchronized, and (b) all the segments are well calibrated (see the next two subsections).

2.4.3 Local synchronization procedure

It may occur that those constrained empirical energy levels which were determined only by very few transitions during the blockMARVEL refinement must be replaced with more dependable energy values using the gsCD relations of the energy levels. Therefore, at this point of our extMARVEL algorithm, an automated analysis, called *local synchronization procedure*, is executed to (a) synchronize, if necessary, the energy levels with their gsCD relations (*i.e.*, substitute the constrained empirical energy levels with higher-quality energy values, derived from those lines of the corresponding gsCDs which are in best agreement), and (b) exclude those transitions whose recalculated fitting defects are positive. The stages of this algorithm are detailed in Appendix C. If there is no energy level which should be synchronized with its gsCD relations, then the SN is said to be *perfectly synchronized*. If our SN is not synchronized perfectly, the blockMARVEL procedure has to be repeated (only once more) upon completion of the local synchronization process.

Table 2 $H_2{}^{16}O$ data source segments employed during this work and some of their most important characteristics.[a]

Segment tag	Range/cm^{-1}	A/V/E	ESU/cm^{-1}	RSU/cm^{-1}	ASU/cm^{-1}	MSU/cm^{-1}
69Kukolich[47]	0.74–0.74	1/1/1	2.000E-09	2.000E-09	2.000E-09	2.000E-09
09CaPuHaGa[57]	10.71–20.70	7/7/7	2.000E-08	2.000E-08	2.000E-08	2.000E-08
71Huiszoon[52]	6.11–6.11	1/1/1	4.000E-08	4.000E-08	4.000E-08	4.000E-08
95MaKr[58]	18.58–18.58	1/1/0	7.000E-08	7.000E-08	9.900E-08	9.900E-08
06GoMaGuKn[53]	6.11–18.58	12/12/11	2.000E-07	1.145E-07	8.788E-08	2.679E-07
18KaStCaCa[101]	7164.90–7185.60	8/8/8	4.000E-07	4.000E-07	4.000E-07	4.000E-07
06MaToNaMo[94]	28.69–165.31	104/103/95	4.000E-06	4.000E-07	1.386E-06	1.786E-05
09CaPuBuTa[95]	36.60–53.44	9/9/9	5.000E-07	5.000E-07	1.101E-06	3.497E-06
11Koshelev[63]	25.09–25.09	1/1/1	5.000E-07	5.000E-07	5.000E-07	5.000E-07
83MeLuHe[65]	16.80–32.95	5/5/5	7.000E-07	7.000E-07	9.818E-07	2.109E-06
83HeMeLu[64]	13.01–32.95	7/7/6	7.000E-07	7.000E-07	4.235E-06	2.411E-05
07KoTrGoPa[56]	10.71–12.68	3/3/3	6.000E-07	8.507E-07	8.584E-07	1.725E-06
71StBe[48]	0.74–6.11	3/3/2	9.000E-07	9.721E-07	5.971E-06	1.525E-05
18ChHuTaSu[46]	12621.74–12665.16	6/6/6	1.000E-06	1.000E-06	1.000E-06	1.000E-06
87BaAlAlPo[76]	14.20–19.08	6/6/5	1.000E-06	1.000E-06	4.249E-06	2.025E-05
79HeJoMc_S1[73]	0.07–0.07	1/1/1	1.000E-05	1.000E-06	3.458E-06	3.458E-06
72LuHeCoGo[54]	6.11–25.09	14/14/13	5.000E-06	1.130E-06	1.454E-06	5.427E-06
95MaOdIwTs[59]	18.58–162.44	139/138/135	1.000E-06	1.531E-06	2.504E-06	3.645E-05
11DrYuPeGu[96]	82.86–90.84	26/25/25	6.000E-06	1.840E-06	3.187E-06	8.888E-06
80Kuze[74]	0.40–4.00	5/5/5	2.000E-06	2.000E-06	9.671E-06	2.986E-05
91AmSc[78]	8.25–11.84	5/5/5	2.000E-06	2.000E-06	2.571E-06	3.642E-06
97DeLoInNo[87]	118.32–119.07	5/5/5	2.000E-06	2.000E-06	1.167E-05	3.066E-05
83BuFeKaPo[61]	16.80–21.55	5/5/5	4.000E-06	2.273E-06	2.266E-06	5.064E-06
13YuPeDr[62]	17.69–67.21	182/180/178	5.000E-06	2.502E-06	3.085E-06	1.136E-04
00ChPePiMa[89]	28.05–52.51	17/17/17	8.000E-06	2.854E-06	5.268E-06	1.994E-05
91PeAnHeLu[79]	4.66–19.80	30/30/30	3.000E-06	3.000E-06	4.258E-06	1.667E-05
96Belov[85]	28.05–30.79	5/5/5	3.000E-05	3.825E-06	5.094E-06	1.075E-05
12YuPeDrMa_S1[97]	9.86–90.77	103/102/100	4.000E-06	4.000E-06	9.628E-06	2.428E-04
95Pearson[84]	4.33–17.22	9/8/8	4.000E-06	4.000E-06	4.000E-06	4.000E-06
51Jen[49]	0.74–0.74	1/1/1	2.000E-05	4.006E-06	4.409E-06	4.409E-06
54PoSt[50]	0.74–0.74	1/1/1	2.000E-06	4.673E-06	5.142E-06	5.142E-06
70StSt[60]	18.58–18.58	1/1/0	3.000E-05	6.241E-06	6.887E-06	6.887E-06
96BrMa[86]	5206.31–5396.54	28/28/24	5.000E-05	7.179E-06	3.420E-05	1.232E-04
72FlCaVa[51]	0.74–25.09	7/7/0	2.000E-05	1.265E-05	1.508E-05	3.015E-05
85BrTo[75]	1323.33–1992.65	71/70/68	1.000E-04	1.827E-05	6.028E-05	3.401E-04
91Toth_S1[80]	1072.61–2265.31	740/733/718	6.000E-05	1.938E-05	4.701E-05	4.917E-04
87BeKoPoTr[77]	7.76–19.85	5/5/5	3.000E-05	1.968E-05	2.250E-05	4.747E-05
06JoPaZeCo[93]	1485.13–1486.16	2/2/2	2.000E-05	2.000E-05	2.956E-05	3.912E-05
05HoAnAlPi[91]	212.56–594.95	166/164/60	3.000E-05	3.122E-05	3.407E-05	3.457E-04
93Totha_S1[81]	1316.13–4260.41	585/585/567	8.000E-05	3.325E-05	9.136E-05	9.443E-04
93Tothb_S1[82]	1881.08–4306.72	1076/1076/1064	8.000E-05	3.585E-05	7.996E-05	1.483E-03
54KiGo[55]	6.11–6.11	1/1/0	5.000E-05	3.766E-05	4.437E-05	4.437E-05
05Toth[92]	2926.46–7641.94	1895/1889/1752	8.000E-05	5.833E-05	1.250E-04	2.795E-03
03ZoVa[90]	3010.23–4044.91	469/469/455	1.000E-04	9.449E-05	2.326E-04	2.459E-03
13LuLiWaLi[98]	12573.30–12752.16	73/73/65	3.000E-05	9.686E-05	6.033E-05	5.242E-04
15SiHo[100]	7714.81–7919.88	71/71/62	1.000E-04	1.000E-04	2.838E-04	4.986E-03
95PaHo[83]	177.86–519.59	246/245/5	7.000E-05	1.067E-04	1.819E-04	2.073E-03
97MiTyKeWi[88]	2507.23–4402.77	935/933/205	6.000E-05	1.528E-04	4.514E-04	6.739E-03

2.4.4 Accuracy-based recalibration (ABR)

Having the constrained energy levels derived, each *ill-calibrated source segment*, for which the average of the residuals related to its transitions significantly differs from zero, has to be recalibrated. The sequential steps of our improved recalibration technique, called *accuracy-based recalibration (ABR)*, are presented in Appendix D. Notice that only those segments can be recalibrated with the ABR procedure which contain *unexploited lines*, *i.e.*, transitions not used for the determination of the energy levels.

[a] Tags denote experimental data-source segments used in this study. The column 'Range' indicates the range corresponding to validated wavenumber entries within the experimental linelist. 'A/V/E' is an ordered triple standing for the number of assigned transitions in the source segments (A) and for the number of transitions validated (V) and exploited (E) in this paper. In columns ESU (estimated segment uncertainties), RSU (refined segment uncertainties), ASU (average segment uncertainties), and MSU (maximum segment uncertainties), "2.000E-09", *e.g.*, stands for 2.000×10^{-9}. Rows of this table are arranged in the order of the RSUs.

2.4.5 Full network characterization (FNC)

As the closing step of the intMARVEL procedure, the energy levels, the transitions, and the segments are given quality labels *via* the so-called *full network characterization (FNC) technique*. During this qualification, described in detail in Appendix E, (a) the energy levels are equipped with uncertainties, dependability grades, and useful connectivity parameters, (b) the transitions are rated using the grades of their upper and lower energy levels, and (c) segments are associated with statistical data concerning their wavenumber ranges and their accuracy.

3 Experimental data sources

An almost complete set of data sources, available up to 2013, about recorded and assigned high-resolution spectra of $H_2{}^{16}O$ is given in the IUPAC compilation, Ref. 17. Since then, partly helped by the data of Ref. 17, results of a number of new spectroscopic measurements and a number of new analyses have been published on $H_2{}^{16}O$, some highly relevant for this study.[45,46,62,98–102] Note that a complete update of the IUPAC Part III paper[17] is in progress.[107]

In Table 2 segments of experimental sources, old and new, are collected reporting highly accurate measured transitions, defined as having ESU $\leq 10^{-4}$ cm^{-1}. These are the segments utilized during this study. For the experimental data sources we follow the IUPAC naming convention.[15] The representation of the segment accuracies can be found in Table 2. It is comfortable to note that good agreement is seen among the ESU, RSU, ASU, and MSU values.

There is only one source reporting hyperfine transitions characterizing the spectra of *ortho*-$H_2{}^{16}O$, 09CaPuHaGa.[57] Nevertheless, not the hyperfine split transitions but their weighted averages, also reported in 09CaPuHaGa,[57] have been utilized in this study.

Whenever possible, the available experimental information was used to determine the ESU values of the source segments. In some cases, when the accuracy of the source segment was not explicitly given, an educated guess had to be employed for the ESU value. When line-by-line uncertainties were reported, *e.g.*, in the cases of 95MaOdIwTs,[59] 12YuPeDrMa,[97] 11DrYuPeGu,[96] 13YuPeDr,[62] and 06MaToNaMo,[94] we formed an average from these data. The renewed MARVEL protocol requires that less accurate segments of sources are ignored from the analysis of more accurate data. In the case of the sources 93Toth,[108] 93Tothb,[82] and 91Toth,[80] only the transitions given with five digits after the decimal point were included in our initial dataset.

Our refinements indicated that a couple of sources, namely 82KaJoHo,[109] 17MoMiKaBe,[45] 18MiMoKaKa,[102] and 14ReOuMiWa,[99] are not sufficiently accurate for the purposes of the present study. Therefore, the transitions these sources contain were omitted from our analysis.

4 Results and discussion

Let us start by providing some statistical information about the present analysis of the $H_2{}^{16}O$ measured transitions based on the highly accurate data source segments reported in Table 2. The number of transition entries in the leading subnetwork is 3087 and 3980 for *para*- and *ortho*-$H_2{}^{16}O$, respectively. As to transitions of A^+/A^- grade (the detailed definition of the grades of the empirical energy levels determined in this study is given in Appendix E), their number is 650/461 for *para*/*ortho*-$H_2{}^{16}O$. The number of lines which had to be excluded from our extMARVEL analysis from the source segments selected is only 26, six of which are in the THz region (*vide infra*). For *para*-$H_2{}^{16}O$, the number of energy levels determined is 725, out of which 97 have A^+/A^- grades. For *ortho*-$H_2{}^{16}O$, the extMARVEL analysis involved 857 energy levels, of which 117 received A^+/A^- grades. The largest

J value, J_{max}, is 17/18 for *para*/*ortho*-$H_2{}^{16}O$. The range of highly accurate $H_2{}^{16}O$ energy levels is quite extended, it is 0–3000 cm^{-1} for *para*-$H_2{}^{16}O$ and somewhat narrower, 0–1800 cm^{-1}, for *ortho*-$H_2{}^{16}O$.

One of the important practical results of this study, a set of highly accurate empirical energy levels of $H_2{}^{16}O$, is reported in Table 3. Only empirical energy levels of grade A^+ and A^- quality are included in Table 3. The highly accurate rovibrational energy levels belong to the vibrational states $(v_1\ v_2\ v_3) = (0\ 0\ 0)$, $(0\ 1\ 0)$, and $(0\ 2\ 0)$, where v_1, v_2, and v_3 stand for the symmetric stretch, bend, and antisymmetric stretch vibrational quantum numbers. All energy levels reported have at least an 8-digit accuracy, often considerably better (up to 10 digits of accuracy). J_{max} is 13/14 for the most dependable rovibrational energy levels presented in Table 3.

4.1 Highly accurate sources

Figure 2 shows the unsigned residuals (differences between the observed and the calculated wavenumbers) for those segments which have RSU $< 10^{-5}$ cm^{-1}. Clearly, the extMARVEL refinement retained the high accuracy of the original measurements.

Some of the extMARVEL rovibrational energy levels have an accuracy considerably lower than that of the transitions determining it. The simplest example that highlights the difficulties of providing uncertainties to energy levels is given by two separate 4-cycles, both formed by highly accurate transitions, connected by a bridge of lower measured accuracy. If one of the 4-cycles contains a core of the spectroscopic network, then all its energy levels have the same high accuracy as the measured transitions. However, this is not true for the other 4-cycle. Compared to the core of this cycle the transitions are known with high accuracy, but the overall accuracy of the energy levels is determined by the accuracy of the bridge. This also means that if the transitions are reconstructed from MARVEL energy levels, they may have considerably higher uncertainties than the directly measured transitions. This characteristics of the MARVEL protocol cannot be easily circumvented without new, accurate measurements. Thus, in MARVEL determinations of empirical rovibronic energy levels it can happen that the reconstructed line frequencies should have a better uncertainty than indicated by the energy-level-based uncertainties. For the present study this means that uncertainties of transitions determined from uncertainties of rovibrational energy levels may not be fully realistic, they may provide exaggerated uncertainties, which must be considered when these transitions are used in an application.

For *para*-$H_2{}^{16}O$ one of the most important lines is the $5_{1,5} \leftarrow 4_{2,2}$ pure rotational transition at about 325.153 GHz (this is how this transition is usually reported) on the (0 0 0) ground vibrational state. Our final 8-digit-accuracy determination gives 325 152 900(10) kHz, based basically upon the experimental result of 06GoMaGuKn,[53] but including the effects of all other measured transitions of this line.

Next, let us turn our attention to the THz region. In the THz region the following transitions had to be excluded from the extMARVEL analysis (all wavenumbers in cm^{-1}): $-36.310\ 4(97)$[62], $-39.003\ 01(58)$[97], $-61.792\ 93(68)$[62], $-82.638\ 9(21)$[59], $-86.467\ 2(10)$[96], and $-118.951\ 4(1)$[94]. 95MaOdIwTs[59] and 06MaToNaMo[94] are two sources which report accurately measured purely rotational transitions in the THz region. The absolute residuals, *i.e.*, the unsigned differences of the observed and the extMARVEL-predicted wavenumbers, are plotted in Fig. 3. The estimated ESU of the transitions in 95MaOdIwTs[59] is 1×10^{-6} cm^{-1}, *i.e.*, 30 kHz, while the RSU, according to our study, is slightly higher, 46 kHz. Figure 3 shows that extMARVEL is basically able to confirm the claimed accuracy

Table 3 Accurate empirical (extMARVEL) energy levels, and their assignments, for *para*-$H_2{}^{16}O$ (first four columns) and *ortho*-$H_2{}^{16}O$ (last four columns), of grade A^+ and A^- quality.

Energy/cm^{-1}	$(v_1\ v_2\ v_3)J_{K_a,K_c}$	Energy/cm^{-1}	$(v_1\ v_2\ v_3)J_{K_a,K_c}$	Energy/cm^{-1}	$(v_1\ v_2\ v_3)J_{K_a,K_c}$	Energy/cm^{-1}	$(v_1\ v_2\ v_3)J_{K_a,K_c}$
0	$(0\,0\,0)0_{0,0}$	1739.4837116(72)	$(0\,1\,0)3_{1,3}$	0	$(0\,0\,0)1_{0,1}$	1718.511395(14)	$(0\,1\,0)2_{2,1}$
37.13712384(73)	$(0\,0\,0)1_{1,1}$	1743.486343(15)	$(0\,1\,0)2_{2,0}$	18.5773849(12)	$(0\,0\,0)1_{1,0}$	1748.619103(14)	$(0\,1\,0)3_{1,2}$
70.0908134(17)	$(0\,0\,0)2_{0,2}$	1774.616260(34)	$(0\,0\,0)12_{1,11}$	55.7020292(20)	$(0\,0\,0)2_{1,2}$	1750.956732(27)	$(0\,0\,0)12_{2,11}$
95.1759380(16)	$(0\,0\,0)2_{1,1}$	1806.671527(29)	$(0\,0\,0)13_{1,13}$	111.1072826(14)	$(0\,0\,0)2_{2,1}$	1765.248467(11)	$(0\,0\,0)8_{8,1}$
136.1639195(14)	$(0\,0\,0)2_{2,0}$	1810.583275(15)	$(0\,0\,0)9_{-,3}$	112.9673030(17)	$(0\,0\,0)3_{0,3}$	1782.875642(30)	$(0\,0\,0)13_{0,13}$
142.2784859(17)	$(0\,0\,0)3_{1,3}$	1813.7875991(80)	$(0\,1\,0)3_{2,2}$	149.57145278(95)	$(0\,0\,0)3_{1,2}$	1786.7935598(19)	$(0\,0\,0)9_{7,2}$
206.3014280(31)	$(0\,0\,0)3_{2,2}$	1817.4511918(64)	$(0\,1\,0)4_{0,4}$	188.36200832(45)	$(0\,0\,0)3_{2,1}$	1789.4290336(18)	$(0\,0\,0)11_{3,8}$
222.0527607(37)	$(0\,0\,0)4_{0,4}$	1843.029600(19)	$(0\,0\,0)11_{4,8}$	201.0440271(10)	$(0\,0\,0)4_{1,4}$	1795.540756(16)	$(0\,1\,0)3_{2,1}$
275.4970429(27)	$(0\,0\,0)4_{1,3}$	1875.461816(14)	$(0\,0\,0)10_{6,4}$	261.6242164(22)	$(0\,0\,0)3_{3,0}$	1797.8024553(79)	$(0\,1\,0)4_{2,3}$
285.2193412(44)	$(0\,0\,0)3_{3,1}$	1875.4697166(71)	$(0\,1\,0)4_{1,3}$	276.567924(18)	$(0\,0\,0)4_{2,3}$	1851.1786135(16)	$(0\,0\,0)10_{0,5}$
315.7795348(13)	$(0\,0\,0)4_{2,2}$	1907.451419(18)	$(0\,1\,0)3_{3,1}$	301.5535454(17)	$(0\,0\,0)5_{0,5}$	1875.2138010(18)	$(0\,0\,0)11_{4,7}$
326.6254684(78)	$(0\,0\,0)5_{1,5}$	1922.829071(26)	$(0\,1\,0)5_{1,5}$	358.7225273(24)	$(0\,0\,0)4_{3,2}$	1883.821403(20)	$(0\,1\,0)3_{3,0}$
383.8425157(20)	$(0\,0\,0)4_{3,1}$	1922.901124(15)	$(0\,1\,0)4_{2,2}$	375.6631531(16)	$(0\,0\,0)5_{1,4}$	1884.2219798(83)	$(0\,1\,0)4_{2,3}$
416.2087412(12)	$(0\,0\,0)5_{2,4}$	1960.207408(42)	$(0\,0\,0)12_{2,10}$	422.7163073(29)	$(0\,0\,0)5_{2,3}$	1896.9721837(78)	$(0\,1\,0)5_{0,5}$
446.6965909(23)	$(0\,0\,0)6_{0,6}$	1985.784890(12)	$(0\,0\,0)11_{5,7}$	423.4579897(25)	$(0\,0\,0)6_{1,6}$	1938.7125366(44)	$(0\,0\,0)12_{3,10}$
488.134180(21)	$(0\,0\,0)4_{4,0}$	2005.917049(24)	$(0\,1\,0)4_{3,1}$	464.3133394(39)	$(0\,0\,0)4_{4,1}$	1975.200945(13)	$(0\,0\,0)11_{5,6}$
503.9681044(18)	$(0\,0\,0)5_{3,3}$	2024.152656(15)	$(0\,1\,0)5_{2,4}$	485.0176947(20)	$(0\,0\,0)5_{3,2}$	1977.068672(11)	$(0\,1\,0)5_{1,4}$
542.9057791(29)	$(0\,0\,0)6_{1,5}$	2041.780525(32)	$(0\,1\,0)6_{0,6}$	529.1170284(28)	$(0\,0\,0)6_{2,5}$	1981.0213294(89)	$(0\,1\,0)4_{3,2}$
586.4791840(34)	$(0\,0\,0)7_{1,7}$	2042.374093(41)	$(0\,0\,0)13_{2,12}$	562.4491822(25)	$(0\,0\,0)7_{0,7}$	1986.010725(17)	$(0\,0\,0)9_{8,1}$
602.7734941(12)	$(0\,0\,0)6_{2,4}$	2054.368662(49)	$(0\,0\,0)10_{7,3}$	586.546807(30)	$(0\,0\,0)5_{4,1}$	2018.516213(38)	$(0\,0\,0)13_{1,12}$
610.1144322(34)	$(0\,0\,0)5_{4,2}$	2105.867903(31)	$(0\,0\,0)12_{3,9}$	625.1843305(33)	$(0\,0\,0)5_{3,4}$	2018.958904(55)	$(0\,1\,0)6_{1,6}$
661.5489140(25)	$(0\,0\,0)6_{3,3}$	2126.407726(17)	$(0\,1\,0)5_{3,3}$	680.4196886(19)	$(0\,0\,0)7_{1,6}$	2030.174340(10)	$(0\,1\,0)5_{2,3}$
709.6082135(35)	$(0\,0\,0)7_{2,6}$	2129.618680(19)	$(0\,1\,0)4_{4,0}$	718.2819234(73)	$(0\,0\,0)5_{5,0}$	2030.550847(17)	$(0\,0\,0)10_{7,4}$
742.073025(23)	$(0\,0\,0)5_{3,1}$	2142.597656(50)	$(0\,0\,0)11_{6,6}$	720.3683039(44)	$(0\,0\,0)8_{1,8}$	2101.157029(29)	$(0\,0\,0)12_{4,9}$
744.063662(18)	$(0\,0\,0)8_{0,8}$	2146.263727(12)	$(0\,0\,0)6_{4,3}$	732.9304170(85)	$(0\,0\,0)6_{4,2}$	2105.804890(30)	$(0\,1\,0)4_{4,1}$
757.7801884(17)	$(0\,0\,0)6_{4,2}$	2205.652745(34)	$(0\,0\,0)12_{4,8}$	758.6154643(26)	$(0\,0\,0)7_{2,5}$	2106.699940(15)	$(0\,1\,0)5_{3,2}$
816.694236(18)	$(0\,0\,0)7_{3,5}$	2211.190640(11)	$(0\,1\,0)6_{2,4}$	818.5622284(16)	$(0\,0\,0)7_{3,4}$	2120.251897(38)	$(0\,0\,0)11_{6,5}$
882.8903279(36)	$(0\,0\,0)8_{1,7}$	2248.064560(40)	$(0\,0\,0)13_{3,11}$	861.805846(20)	$(0\,0\,0)8_{2,7}$	2137.4916804(79)	$(0\,1\,0)6_{3,5}$
888.6326507(25)	$(0\,0\,0)6_{5,1}$	2300.685031(57)	$(0\,0\,0)12_{5,7}$	864.804378(27)	$(0\,0\,0)6_{5,2}$	2223.090464(31)	$(0\,0\,0)13_{2,11}$
920.210001(21)	$(0\,0\,0)9_{1,9}$	2321.812959(39)	$(0\,0\,0)11_{7,5}$	896.373981(28)	$(0\,0\,0)9_{0,9}$	2228.068156(20)	$(0\,1\,0)5_{4,1}$
927.7439024(25)	$(0\,0\,0)7_{4,4}$	2327.883768(31)	$(0\,0\,0)14_{1,13}$	907.4427403(43)	$(0\,0\,0)7_{4,3}$	2230.489505(29)	$(0\,0\,0)10_{8,3}$
982.9117144(33)	$(0\,0\,0)8_{2,6}$	2399.165485(19)	$(0\,1\,0)6_{3,2}$	982.3215733(24)	$(0\,0\,0)8_{3,6}$	2247.917879(13)	$(0\,1\,0)6_{3,4}$
1045.058341(10)	$(0\,0\,0)6_{6,0}$	2426.196189(33)	$(0\,0\,0)13_{4,10}$	1021.2635825(52)	$(0\,0\,0)6_{6,1}$	2251.578499(37)	$(0\,0\,0)12_{5,8}$
1050.1576637(25)	$(0\,0\,0)8_{3,5}$	2522.261315(75)	$(0\,0\,0)11^{+}_{8,4}$	1036.04108419(92)	$(0\,0\,0)7_{5,2}$	2285.9358498(85)	$(0\,1\,0)7_{1,6}$
1059.646655(16)	$(0\,0\,0)7_{5,3}$	2613.104557(29)	$(0\,0\,0)12^{-}_{7,5}$	1055.2852146(29)	$(0\,0\,0)9_{1,8}$	2298.111387(31)	$(0\,0\,0)11_{7,4}$
1080.3854445(49)	$(0\,0\,0)9_{2,8}$	2670.789697(11)	$(0\,1\,0)8_{1,5}$	1090.755548(37)	$(0\,0\,0)10_{1,10}$	2304.119655(41)	$(0\,0\,0)14_{2,13}$
1114.532254(42)	$(0\,0\,0)10_{0,10}$	2748.099742(28)	$(0\,0\,0)13^{+}_{6,8}$	1098.9141697(26)	$(0\,0\,0)8_{4,5}$	2313.872429(22)	$(0\,1\,0)8_{1,8}$
1131.7755717(28)	$(0\,0\,0)8_{4,4}$	2920.132079(16)	$(0\,1\,0)8_{5,3}$	1178.1271344(26)	$(0\,0\,0)9_{2,7}$	2368.798205(12)	$(0\,1\,0)7_{2,5}$
1216.189769(16)	$(0\,0\,0)7_{6,2}$			1192.400139(15)	$(0\,0\,0)7_{6,1}$	2374.587137(29)	$(0\,1\,0)6_{4,3}$
1216.2312587(35)	$(0\,0\,0)9_{3,7}$			1231.3723853(22)	$(0\,0\,0)8_{5,4}$	2390.929065(28)	$(0\,0\,0)13_{3,10}$
1255.9115449(16)	$(0\,0\,0)8_{4,3}$			1259.12474290(81)	$(0\,0\,0)9_{5,6}$	2410.006048(34)	$(0\,0\,0)12_{6,7}$
1293.0181387(64)	$(0\,0\,0)10_{2,9}$			1269.8396781(381)	$(0\,0\,0)10_{2,9}$	2439.0809331(76)	$(0\,1\,0)7_{3,4}$
1327.117600(26)	$(0\,0\,0)11_{1,11}$			1303.315623(40)	$(0\,0\,0)11_{0,11}$	2471.3715092(74)	$(0\,1\,0)8_{2,7}$
1340.8848754(28)	$(0\,0\,0)9_{4,6}$			1336.440969(16)	$(0\,0\,0)9_{4,5}$	2498.470793(41)	$(0\,0\,0)11_{8,3}$
1394.8141601(62)	$(0\,0\,0)7_{7,1}$			1371.019836(11)	$(0\,0\,0)7_{7,0}$	2509.998802(13)	$(0\,0\,0)13_{4,9}$
1411.6418912(23)	$(0\,0\,0)8_{6,2}$			1387.8170742(24)	$(0\,0\,0)8_{6,3}$	2527.689135(16)	$(0\,0\,0)14_{3,12}$
1437.9685810(38)	$(0\,0\,0)10^{-}_{8,8}$			1422.333875(16)	$(0\,0\,0)10_{3,8}$	2548.3448233(64)	$(0\,1\,0)7_{4,3}$
1474.9807818(16)	$(0\,0\,0)9_{5,5}$			1453.5030009(27)	$(0\,0\,0)9_{5,4}$	2589.005437(27)	$(0\,0\,0)12_{7,6}$
1525.135986(34)	$(0\,0\,0)11^{-}_{2,10}$			1501.053523(39)	$(0\,0\,0)11_{1,10}$	2605.540083(36)	$(0\,0\,0)13_{5,8}$
1538.149472(38)	$(0\,0\,0)10_{3,7}$			1534.0533340(28)	$(0\,0\,0)12_{1,12}$	2606.3982567(69)	$(0\,1\,0)8_{3,6}$
1557.844416(23)	$(0\,0\,0)12_{0,12}$			1557.541614(23)	$(0\,0\,0)10_{4,7}$	2607.474505(36)	$(0\,0\,0)15_{1,14}$
1590.69070(13)	$(0\,0\,0)8_{7,1}$			1566.895703(55)	$(0\,0\,0)8^{-}_{7,2}$	2664.285572(28)	$(0\,1\,0)9^{-}_{1,8}$
1616.453049(39)	$(0\,0\,0)10_{4,6}$			1594.762770(17)	$(0\,1\,0)10_{0,1}$	2700.3727805(67)	$(0\,1\,0)7_{5,2}$
1631.245482(32)	$(0\,0\,0)9_{6,4}$			1607.588646(24)	$(0\,0\,0)9_{6,3}$	2794.603744(36)	$(0\,1\,0)9_{2,7}$
1634.9670937(84)	$(0\,1\,0)1_{1,1}$			1616.711482(38)	$(0\,1\,0)11_{0}$	2894.450619(52)	$(0\,0\,0)14^{+}_{5,10}$
1664.964584(12)	$(0\,1\,0)2_{0,2}$			1653.267094(16)	$(0\,1\,0)2_{1,2}$	2895.8386504(46)	$(0\,1\,0)8_{5,4}$
1693.649942(12)	$(0\,1\,0)2_{1,1}$			1666.870013(19)	$(0\,0\,0)11_{2,9}$	3368.954986(20)	$(0\,2\,0)3_{2,1}$
1695.068592(55)	$(0\,0\,0)11_{3,9}$			1694.9244456(25)	$(0\,0\,0)10_{5,6}$	4383.251956(15)	$(0\,2\,0)6_{6,1}$
1724.705391(17)	$(0\,0\,0)10_{5,5}$			1708.102334(19)	$(0\,1\,0)3_{0,3}$		

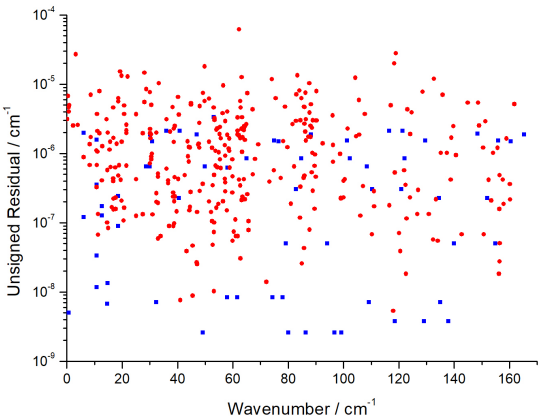


Fig. 2 Unsigned residuals (differences between the observed and the calculated wavenumbers) for the source segments with RSU $\approx 10^{-7}$ cm^{-1} (blue dots), where RSU stands for "refined segment uncertainty", and RSU $\approx 10^{-6}$ cm^{-1} (red dots). Residuals below 10^{-9} cm^{-1} are not plotted as these are artificial results.

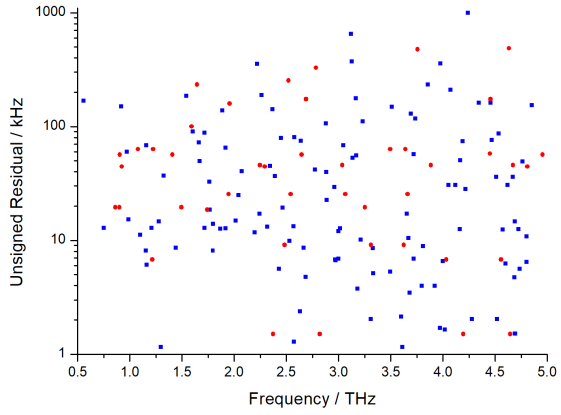


Fig. 3 Unsigned residuals for 95MaOdIwTs[59] and 06MaToNaMo,[94] denoted with blue and red dots, respectively. Residuals below 1 kHz are not plotted as these are artificial results.

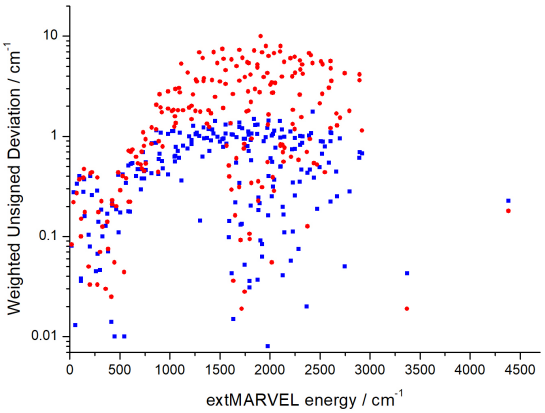


Fig. 5 Unsigned deviations [see Eq. (14)] of the extMARVEL (this work) and the IUPAC Part III[17] energy levels, denoted with blue and red points, respectively, from those obtained in 01LaCoCa[44]).

of the measurements when they are refined together with all the other sources. Only very few transitions have absolute residuals higher than 100 kHz. This means that the extMARVEL procedure is able to retain the accuracy of the original measurements.

It should prove beneficial for future studies to know more and more highly accurate transitions in a wider and wider spectral range. For example, for *ortho*-$H_2{}^{16}O$, two often quoted pure rotational transitions are 321.23 GHz for $10_{2,9} \leftarrow 9_{3,6}$ on (0 0 0) and 336.23 GHz for $5_{2,3} \leftarrow 6_{1,6}$ on (0 1 0). The extMARVEL counterpart of the first transition determined in this study is 321 225 677(13) kHz, while the frequency of the second transition is determined to be 336 228 014(41) kHz, both of which can be considered dependable due to the fact that they are based on energy levels of A^+ grade.

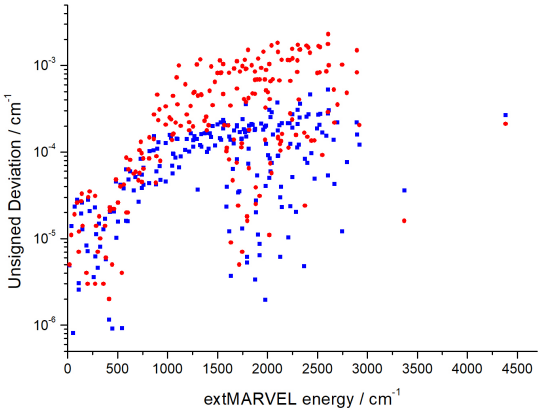


Fig. 4 Unsigned deviations of the extMARVEL and IUPAC Part III energy level sets, denoted with blue and red dots, respectively, referenced to those determined by Lanquetin *et al.*[44]

4.2 IAU lines

For all 13 $H_2{}^{16}O$ lines recommended by the International Astronomical Union (IAU) empirical estimates have been derived in this study. These estimates are listed in Table 1. Note that all these IAU selected frequencies are below 1 THz. This is due partly to the fact that when these lines were selected THz spectroscopy measurements were rarely available.

The data presented in Table 1 allows an assessment of the extMARVEL frequencies in relation to those deemed to be most important by IAU. As evident from the last column of Table 1, all these lines have been determined experimentally via multiple, independent, highly accurate measurements, contributing to an improved confidence in the astronomically important frequencies.

4.3 Comparison with the energy levels of Lanquetin *et al.*

Figure 4 shows the unsigned differences of the extMARVEL and the IUPAC energy levels with respect to the JPL data, which are the same as those given by Lanquetin *et al.*[44]

The first observation one can make about Fig. 4 is that the extMARVEL protocol is a significant improvement over the standard MARVEL protocol employed to obtain the IUPAC Part III energy levels.[17] It is expected that the relation

$$|\bar{E}_{01\mathrm{LaCoCa},i} - \bar{E}_{M,i}| \leq \sqrt{\varepsilon^2_{01\mathrm{LaCoCa},i} + \varepsilon^2_{M,i}} \qquad (13)$$

should be (approximately) satisfied, where (a) $\bar{E}_{M,i}$ is the energy of the ith energy level obtained in this paper (M = extMARVEL) and in the IUPAC Part III paper[17] (M = IUPAC) with $\varepsilon_{M,i}$ uncertainty, and (b) $\bar{E}_{01\mathrm{LaCoCa},i}$ is the counterpart of $\bar{E}_{M,i}$ estimated in the source 01LaCoCa[44] with an $\varepsilon_{01\mathrm{LaCoCa},i}$ uncertainty. As a result, the corresponding *weighted unsigned deviation* (WUD), defined as

$$D^{[\mathrm{w}]}_{M,i} = \frac{|\bar{E}_{01\mathrm{LaCoCa},i} - \bar{E}_{M,i}|}{\sqrt{\varepsilon^2_{01\mathrm{LaCoCa},i} + \varepsilon^2_{M,i}}}, \qquad (14)$$

must not be larger than 1. These weighted unsigned deviations

are plotted in Fig. 5.

The substantial improvement achieved during the present study as compared to the IUPAC Part III study is obvious from Fig. 5. The relatively large WUDs related to the IUPAC Part III data (in fact, the assigned uncertainties are smaller, by an order of magnitude, than the corresponding unsigned deviations), are attributed to the inaccuracy and the too optimistic energy uncertainties of some of the IUPAC Part III energy levels.

It is also worth examining whether the conditions

$$|\bar{E}_{\mathrm{01LaCoCa},i} - \bar{E}_{\mathrm{extMARVEL},i}| \leq \varepsilon_{\mathrm{extMARVEL},i} \tag{15}$$

are (approximately) satisfied. Accordingly, the *relative unsigned deviations* (RUD),

$$D_i^{[\mathrm{r}]} = \frac{|\bar{E}_{\mathrm{01LaCoCa},i} - \bar{E}_{\mathrm{extMARVEL},i}|}{\varepsilon_{\mathrm{extMARVEL},i}} \leq 1, \tag{16}$$

are formed and illustrated in Fig. 6. When these deviations are larger than 1, the 01LaCoCa[44] energies are not within the $\bar{E}_{\mathrm{extMARVEL},i} \pm \varepsilon_{\mathrm{extMARVEL},i}$ intervals. The protocol used in this study for the calculation of the uncertainties of the rovibrational energy levels is reliable; thus, relative deviations greater than 1 indicate that our energy levels are more accurate than those derived by Lanquetin *et al.*,[44] as also supported by the data of Table 1.

Figure 7 shows the 01LaCoCa[44] uncertainties relative to their extMARVEL counterparts. Clearly, the present study represents a significant improvement over the uncertainties of the energies given in the source 01LaCoCa.[44]

Finally, let us support the conclusions based on the figures presented with some statistical data, including unsigned (UD), weighted (WUD), and relative unsigned (RUD) deviations, all concerning a comparison with the data of Lanquetin *et al.* ,[44] handling them as reference data. The average and the maximum UD decreased from 4.4×10^{-4} and 2.3×10^{-3} to 1.1×10^{-4} and 5.2×10^{-4} cm^{-1}, respectively. The average and the maximum WUD decreased from 2.21 to 0.59 and from 9.99 to 1.76, respectively. These substantial decreases in the deviations clearly prove that the extMARVEL treatment allows the full utilization of the most accurate spectroscopic measurements, as now we exceed the internal accuracy of the data presented by Lanquetin *et al.*[44]

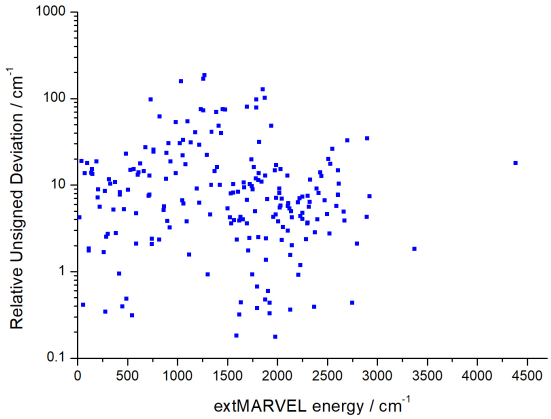


Fig. 6 Relative unsigned deviations [see Eq. (16)] of the energy levels found in 01LaCoCa[44] from the extMARVEL energy levels.

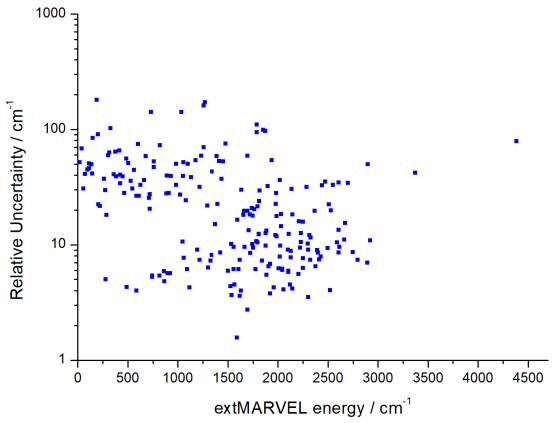


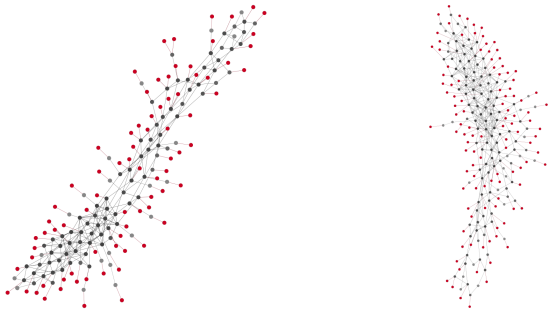
Fig. 7 Uncertainties of the the energy levels in 01LaCoCa[44] relative to their extMARVEL counterparts.

5 Conclusions

During this study the standard Measured Active Rotational-Vibrational Energy Levels (MARVEL) algorithm[11,12] has been extended, to improve its performance toward allowing the automatic determination of highly accurate empirical energy levels, matching or exceeding the accuracy of the best underlying spectroscopic measurements. This is an especially important undertaking in the era of optical-frequency-comb spectroscopies, as they can consistently yield orders of magnitude more accurate transitions than traditional high-resolution spectroscopic techniques.

There are several important algorithmic changes introduced in this study into what we call the extended MARVEL protocol. It is worth reiterating these improvements one by one as they all contributed to increase the utility of the MARVEL analysis of experimental transitions data.

First, unlike in the standard version, in the new, extended protocol MARVEL-type analyses are performed based on the use of groups of transitions blocked by their estimated experimental uncertainties. This requires that the user segments the input sources based on assumed uncertainties of different groups of transitions (a line-by-line analysis yielding individual initial line uncertainties would be ideal but this appears to be too demanding). Second, the inversion and weighted least-squares refinement procedure is now based on sequential addition of blocks of decreasing accuracy. Energy levels determined in a given block are not allowed to be changed by less accurate measurements. Third, spectroscopic cycles are introduced during the refinement process. This is a particularly important advancement as due to the law of energy conservation[13] one can detect straightforwardly the best as well as the worst transitions in the collated set of



(a) *para*-$H_2{}^{16}O$ **(b)** *ortho*-$H_2{}^{16}O$

Fig. 8 The *para*- and *ortho*-$H_2{}^{16}O$ principal components of the experimental spectroscopic network of highly accurate measured transitions, including only those energy levels and transitions which have a grade of at least A$^-$ quality. Note that the energy levels were determined by a much larger number of transitions, all those which have an uncertainty of better than 10^{-4} cm^{-1}.

experimentally measured and assigned transitions. Fourth, automated recalibration of the segments requiring this adjustment is performed. As shown before,[17] MARVEL is able to perform this job quite reliably. Fifth, synchronization of the ground-state combination difference relations is performed to reduce residual uncertainties in the resulting dataset of empirical (MARVEL) energy levels. Sixth, an improved classification scheme, providing seven grades decreasing in quality from A$^+$ to D, of the empirical energy levels is introduced, the grading is based on the assumed accuracy and dependability of the energy levels. Energy levels can obtain the highest grade only if they are part of several dependable cycles and are involved in a number of accurate measurements. This grading of the energy levels automatically results in a grading of the measured transitions, as the lower grade of the lower and the upper energy level is attached to the transition they define.

We used $H_2{}^{16}O$ as the molecule of choice for our feasibility study testing the extended MARVEL protocol. Since the International Astronomical Union selected 13 water lines as highly important for astrophysical applications, it is important to note that all these transitions are reproduced now perfectly well by the extMARVEL protocol. From an application point of view it is also important to note that all these transitions are measured extremely accurately, though they are not part of cycles of similar high accuracy. This calls for new measurements, most likely involving optical-frequency-comb spectroscopic techniques. For *para*- and *ortho*-$H_2{}^{16}O$, we determined 97 and 117 energy levels with grades of at least A$^-$, meaning an accuracy better than 10^{-4} cm^{-1}. The range covered by these highly accurate rovibrational energies is quite substantial, 0–3000 cm^{-1} for *para*-$H_2{}^{16}O$ and 0–1800 cm^{-1} for *ortho*-$H_2{}^{16}O$.

The present dataset of highly accurate energy levels is larger than the dataset maintained at JPL, which is the same as the one published by Lanquetin *et al.*[44] Furthermore, on average our data have an accuracy about an order of magnitude better than the data of Lanquetin *et al.*[44] Note also that the highly accurate energy levels are part of a large number of cycles, as shown in Fig. 8.

Appendix A Stages of the HSP procedure (see Section 2.4.1)

S.1 for all $1 \le i \le N_{\mathfrak{s}}$: set $\mathrm{RSU}^{[\mathfrak{s}(i)]} = \mathrm{ESU}^{[\mathfrak{s}(i)]}$, where

(a) $\mathrm{ESU}^{[\mathfrak{s}(i)]}$ is the *estimated segment uncertainty* of the $\mathfrak{s}(i)$ segment,

(b) $\mathrm{RSU}^{[\mathfrak{s}(i)]}$ is the (actual) *refined segment uncertainty* of $\mathfrak{s}(i)$, and

(c) $N_{\mathfrak{s}}$ is the number of segments in the SN;

S.2 for all $1 \le i \le n_{\mathrm{mh}}$, where $n_{\mathrm{mh}}(=10)$ is the *maximum number of HSP iterations*:

S.2.1 save the indices of the segments into $k_1, k_2, \ldots, k_{N_{\mathfrak{s}}}$ so that $\mathrm{RSU}^{[\mathfrak{s}(k_l)]} \le \mathrm{RSU}^{[\mathfrak{s}(k_m)]}$ is enforced for all $1 \le l \le m \le N_{\mathfrak{s}}$;

S.2.2 for all $1 \le j \le N_{\mathfrak{s}}$:

S.2.2.1 set $s = \mathfrak{s}(k_j)$;

S.2.2.2 initialize $\mathrm{RSU}^{[s]}_{\mathrm{prev}}$ as $\mathrm{RSU}^{[s]}$;

S.2.2.3 if $n^{[s]} = 0$, then **continue S.2.2**, where $n^{[s]}$ is the number of rovibrational lines in $\mathscr{N}_{\mathrm{LS}}$ and s;

S.2.2.4 create a $\mathscr{N}_{\mathrm{per}}$ *perturbation subnetwork*[I] with its $\mathbf{P}^{\langle\mathscr{N}_{\mathrm{per}}\rangle}$ participation matrix from $\mathscr{N}_{\mathrm{LS}}$ as follows: $P_m^{\langle\mathscr{N}_{\mathrm{per}}\rangle} = 1$ if $P_m = 1$ and either $\mathfrak{s}(\mathfrak{i}_m) = s$ or $\phi_{\mathrm{per}}\mathrm{RSU}^{[s]} \ge \mathrm{RSU}^{[\mathfrak{s}(\mathfrak{i}_m)]}$, else $P_m^{\langle\mathscr{N}_{\mathrm{per}}\rangle} = 0$ for all $1 \le m \le N_{\mathrm{T}}$, where

(a) $\mathfrak{i}_m$ is the mth *segment index* for which $\mathfrak{s}(\mathfrak{i}_m)$ includes the mth transition, and

(b) $\phi_{\mathrm{per}} \in (0,1]$ is the *perturbation factor* ($\phi_{\mathrm{per}} = 0.1$ is the default value);

S.2.2.5 construct a cycle basis (CB) for $\mathscr{N}_{\mathrm{per}}$, designated with $\mathscr{C}_{\mathrm{B}}$ (perhaps with the *breadth-first search* [BFS] method[110]);

S.2.2.6 calculate the *reduced discrepancy*[II] of every $C \in \mathscr{C}_{\mathrm{B}}$ as

$$\mathrm{RD}_C^{[s]} = \begin{cases} 0, \text{ if } n_C^{[s]} = 0, \\ \left(\mathscr{D}_C - \sum\limits_{\substack{m=1 \\ \mathfrak{s}(m)\neq s}}^{N_{\mathfrak{s}}} n_C^{[\mathfrak{s}(m)]} \mathrm{RSU}^{[\mathfrak{s}(m)]} \right) / n_C^{[s]}, \text{ otherwise,} \end{cases} \tag{17}$$

where

(a) $n_C^{[s]}$ is the number of transitions in C from s, and

(b) $\mathscr{D}_C$ is the discrepancy of C;

S.2.2.7 collect the $C \in \mathscr{C}_{\mathrm{B}}$ *representative cycles*, for which $\mathrm{RD}_C^{[s]} > \phi_{\mathrm{hin}}\mathrm{RSU}^{[s]}$, into the set ρ, where $\phi_{\mathrm{hin}}(=0.01)$[III] is the *reduction hindrance factor*;

S.2.2.8 if $|\rho| = 0$, then **continue S.2.2**, otherwise determine the *average and maximum reduced discrepancy* of the representative cycles for s:

$$\mathrm{ARD}^{[s]} = |\rho|^{-1} \sum\nolimits_{C\in\rho} \mathrm{RD}_C^{[s]}, \tag{18}$$

$$\mathrm{MRD}^{[s]} = \max\nolimits_{C\in\rho} \mathrm{RD}_C^{[s]},$$

where $|\rho|$ is the *cardinality* of the set ρ;

S.2.2.9 if $\mathrm{MRD}^{[s]} > \phi_{\mathrm{ter}}\mathrm{RSU}^{[s]}$, where $\phi_{\mathrm{ter}}(=200)$ is the *termination factor*, then **exit** from the intMARVEL analysis;[V]

S.2.2.10 if $n_{\mathrm{nb}}^{[s]} \ge \lfloor \phi_{\mathrm{rep}} n^{[s]} \rfloor$,[IV] then set $\mathrm{RSU}^{[s]} = \max(\mathrm{ARD}^{[s]}, \phi_{\mathrm{sep}}\mathrm{ESU}^{[s]})$, otherwise use $\mathrm{RSU}^{[s]} = \mathrm{RSU}^{[s]}_{\mathrm{prev}}$, where

(a) $\lfloor\rfloor$ is the *floor operation*;

(b) $n_{\mathrm{nb}}^{[s]}$ is the number of non-bridge lines from s in $\mathscr{N}_{\mathrm{per}}$, and

(c) $\phi_{\mathrm{rep}}(=0.1)$ and $\phi_{\mathrm{sep}}(=0.2)$ are the *representativity* and the *separation factors*, respectively;

S.2.3 if $r_{\max} < \phi_{\mathrm{conv}}$, then **break S.2**, where

(a) $r_{\max} = \max_{j=1}^{N_{\mathfrak{s}}} r_j$,

(b) $r_j = \max(\mathrm{RSU}^{[\mathfrak{s}(j)]}/\mathrm{RSU}^{[\mathfrak{s}(j)]}_{\mathrm{prev}}, \mathrm{RSU}^{[\mathfrak{s}(j)]}_{\mathrm{prev}}/\mathrm{RSU}^{[\mathfrak{s}(j)]})$, and

(c) $\phi_{\text{conv}}(=3)$ is the *convergence factor*;

S.3 for all $1 \le j \le N_{\text{T}}$: if $\text{RSU}^{[\mathfrak{s}(\mathfrak{i}_j)]} = \text{ESU}^{[\mathfrak{s}(\mathfrak{i}_j)]}$, then apply $\delta_{0,j} = \text{RSU}^{[\mathfrak{s}(\mathfrak{i}_j)]}$, otherwise set $\delta_{0,j} = \phi_{\text{us}}\text{RSU}^{[\mathfrak{s}(\mathfrak{i}_j)]}$, where

(a) $\delta_{0,j}$ is the *initial uncertainty* of the jth transition, and

(b) $\phi_{\text{us}}(=0.5)$ is the *uncertainty scaling factor*;

S.4 for all $1 \le j \le N_{\text{T}}$: if $P_j = 1$, then set $B_j = \lfloor \log_{10}(\text{RSU}^{[\mathfrak{s}(\mathfrak{i}_j)]}) \rfloor$, otherwise use $B_j = 100$,[VI] where B_j is the *block index* of the jth line;

S.5 set $B_{\min} = \min_{j=1;P_j=1}^{N_{\text{T}}} B_j$ and $B_{\max} = \max_{j=1;P_j=1}^{N_{\text{T}}} B_j$;

S.6 identify the components of $\mathscr{N}_{\text{LS}}$;

S.7 save the core(j)[VII] and comp(k) indices for all $1 \le j \le N_{\text{c}}$ and $1 \le k \le N_{\text{L}}$, respectively, by means of Eq. (4);

S.8 **end** procedure.

Notes:

[I]The $\mathscr{N}_{\text{per}}$ subnetwork allows the study of the lines arising from s and included in cycles where there is no transition from those s^* segments with $\text{ESU}^{[s^*]} > \text{ESU}^{[s]}$.

[II]The reduced discrepancy of a cycle C related to s represents the inaccuracy of the lines from s in C. Note that nonpositive reduced discrepancies cannot be used for refinement purposes.

[III]In what follows, the default values of the so-called *control parameters*, needed for the intMARVEL procedure, are mostly given in parentheses (see, *e.g.*, '$\phi_{\text{hin}}(=0.01)$' in stage **S.2.2.7**).

[IV]In this case, the prior cleansing should be continued to decrease the reduced discrepancies.

[V]If there are at least $\lfloor \phi_{\text{rep}} n^{[s]} \rfloor$ lines from s in the cycles of $\mathscr{N}_{\text{per}}$, then its RSU is modified by the corresponding ARD value, otherwise this RSU will remain unchanged in the given iteration.

[VI]If the block index of a transition equals 100, this line will not participate in the blockMARVEL refinement (see Appendix B).

[VII]The core indices are determined *via* a traditional MARVEL analysis, using Eq. (4). To accelerate this identification process, a logical variable, u_{ST}, can also be introduced: if $u_{\text{ST}} = 1$, then only lines related to a BFS-based ST of the LS are employed in Eq. (4); otherwise, all the transitions of the LS are utilized in Eq. (4). Obviously, $u_{\text{ST}} = 0$ is a more stable choice than $u_{\text{ST}} = 1$.

Appendix B Stages of the blockMARVEL procedure (see Section 2.4.2)

S.1 for all $1 \le i \le N_{\text{L}}$: set $b_i = 100$, where b_i is the *block index*[I] of the ith energy level;

S.2 choose $\mathscr{N}_{\text{p}}$ as an *empty subnetwork*, where $P_j^{\langle \mathscr{N}_{\text{p}} \rangle} = 0$ for all $1 \le j \le N_{\text{T}}$;

S.3 assign $N_{\text{p,c}} = N_{\text{L}}$;

S.4 for all $1 \le i \le N_{\text{L}}$: set $\text{core}_{\text{p}}(i) = \text{comp}_{\text{p}}(i) = i$ and $\beta_i = 0$;

S.5 for all $1 \le j \le N_{\text{T}}$: initialize δ_j as $\delta_{0,j}$;

S.6 for all $B_{\min} \le \mathcal{B} \le B_{\max}$:

S.6.1 build a $\mathscr{N}_{\text{a}}$ subnetwork of $\mathscr{N}_{\text{LS}}$ in the following fashion: if $B_j \le \mathcal{B}$, then $P_j^{\langle \mathscr{N}_{\text{a}} \rangle} = 1$, otherwise $P_j^{\langle \mathscr{N}_{\text{a}} \rangle} = 0$ for all $1 \le j \le N_{\text{T}}$;

S.6.2 if $\mathbf{P}^{\langle \mathscr{N}_{\text{a}} \rangle} = \mathbf{P}^{\langle \mathscr{N}_{\text{p}} \rangle}$, then **continue S.6**;

S.6.3 find the components of the $\mathscr{N}_{\text{a}}$ subnetwork;

S.6.4 save the $\text{comp}_{\text{a}}(1), \text{comp}_{\text{a}}(2), \ldots, \text{comp}_{\text{a}}(N_{\text{L}})$ component indices of the energy levels;

S.6.5 'infinite' loop:

S.6.5.1 solve Eqs. (11)–(12);[II]

S.6.5.2 determine $\bar{\mathbf{E}}$ using Eq. (9);

S.6.5.3 for all $1 \le j \le N_{\text{T}}$ with $B_j = \mathcal{B}$: calculate Δ_j and d_j;

S.6.5.4 search for $d_{\max} = \max_{j=1;B_j=\mathcal{B}}^{N_{\text{T}}} d_j$;[III]

S.6.5.5 if $d_{\max} \le \phi_{\text{noi}}$, then **break S.6.5**, where $\phi_{\text{noi}}(=10^{-10})$ designates the *numerical noise factor*;

S.6.5.6 for every $1 \le j \le N_{\text{T}}$: if $P_j^{\langle \mathscr{N}_{\text{a}} \rangle} = 1$ and $d_j > \phi_{\text{disc}} d_{\max}$, then substitute δ_j with $\phi_{\text{inc}}|\Delta_j|$, where

(a) $\phi_{\text{disc}}(=0.1)$ is the *discrimination factor*, and

(b) $\phi_{\text{inc}}(=1.1)$ denotes the *increase factor*;

S.6.5.7 create a $\mathscr{N}_{\text{F}}$ *filtered subnetwork* with its $\mathbf{P}^{\langle \mathscr{N}_{\text{F}} \rangle}$ participation matrix in the following fashion: if $P_j^{\langle \mathscr{N}_{\text{a}} \rangle} = 1$ and $\delta_j > \phi_{\text{fil}}\text{RSU}^{[\mathfrak{s}(\mathfrak{i}_j)]}$, then $P_j^{\langle \mathscr{N}_{\text{F}} \rangle} = 1$, otherwise $P_j^{\langle \mathscr{N}_{\text{F}} \rangle} = 0$ for all $1 \le j \le N_{\text{T}}$, where $\phi_{\text{fil}}(=10)$ is the *filtration factor*;

S.6.5.8 save the indices of the transitions in $\mathscr{N}_{\text{F}}$ into $l_1, l_2, \ldots, l_{N_{\text{F}}}$ in the order[IV,V] that either $\lfloor \log_{10}(d_{l_i}) \rfloor > \lfloor \log_{10}(d_{l_j}) \rfloor$ or $\lfloor \log_{10}(d_{l_i}) \rfloor = \lfloor \log_{10}(d_{l_j}) \rfloor$ and $B_{l_i} \ge B_{l_j}$ holds for all $1 \le i \le j \le N_{\text{F}}$, where N_{F} is the number of lines in $\mathscr{N}_{\text{F}}$ (*exclusion sort*);

S.6.5.9 for all $1 \le j \le N_{\text{T}}$: set $P_j^{\langle \mathscr{N}_{\text{a}} \rangle}$ to $P_j^{\langle \mathscr{N}_{\text{a}} \rangle}(1 - P_j^{\langle \mathscr{N}_{\text{F}} \rangle})$;

S.6.5.10 for all $1 \le i \le N_{\text{F}}$:

S.6.5.10.1 identify the components of $\mathscr{N}_{\text{a}}$;

S.6.5.10.2 if the l_ith line connects components in $\mathscr{N}_{\text{a}}$, then reset $P_{l_i}^{\langle \mathscr{N}_{\text{a}} \rangle} = 1$, otherwise set $P_{l_i} = 0$ and $B_{l_i} = 100$;

S.6.5.11 for all $1 \le j \le N_{\text{T}}$: set $P_j^{\langle \mathscr{N}_{\text{F}} \rangle}$ to $(1 - P_j^{\langle \mathscr{N}_{\text{a}} \rangle}) P_j^{\langle \mathscr{N}_{\text{F}} \rangle}$;

S.6.5.12 if $\text{tr}(\mathbf{P}^{\langle \mathscr{N}_{\text{F}} \rangle}) > 0$, then set $\delta_j = \delta_{0,j}$ for every $1 \le j \le N_{\text{T}}$ with $B_j = \mathcal{B}$, where $\text{tr}(\mathbf{P}^{\langle \mathscr{N}_{\text{F}} \rangle})$ is the *trace* of $\mathbf{P}^{\langle \mathscr{N}_{\text{F}} \rangle}$;

S.6.6 for all $1 \le i \le N_{\text{L}}$: if $b_i = 100$ and $\text{core}_{\text{a}}(\text{comp}_{\text{a}}(i)) = \text{core}(\text{comp}(i))$ together with $\mathscr{S}_{\text{a}}(\text{comp}_{\text{a}}(i)) > 1$, then set the b_i index to $\mathcal{B}$, where

(a) $\text{core}_{\text{a}}(j)$ is the jth core index of $\mathscr{N}_{\text{a}}$, and

(b) $\mathscr{S}_{\text{a}}(j)$ is the jth *component size*;[VI]

S.6.7 replace $N_{\text{p,c}}$ with $N_{\text{a,c}}$, where $N_{\text{a,c}}$ is the number of components in $\mathscr{N}_{\text{a}}$;

S.6.8 for all $1 \le i \le N_{\text{p,c}}$: overwrite $\text{core}_{\text{p}}(i)$ with $\text{core}_{\text{a}}(i)$;

S.6.9 for all $1 \le i \le N_{\text{L}}$: set $\text{comp}_{\text{p}}(i) = \text{comp}_{\text{a}}(i)$ and $\beta_i = \bar{E}_i$;

S.7 for all $1 \le j \le N_{\text{T}}$ with $P_j = 0$:

(a) update Δ_j and d_j;

(b) set $\delta_j = \max(\delta_{0,j}, \phi_{\text{inc}}|\Delta_j|)$;

S.8 **end** procedure.

Notes:

[I]At the termination of the blockMARVEL process, the b_i block index denotes the $\mathcal{B}$ value when (a) the ith energy level becomes a non-isolated node in $\mathscr{N}_{\text{a}}$, and (b) the component of this energy level contains at least two vertices, among which at least one core of $\mathscr{N}_{\text{LS}}$ can also be found. In the case that b_i remains 100, the ith energy level became an isolated node of $\mathscr{N}_{\text{LS}}$ during the course of prior cleansing.

[II]The core indices of $\mathscr{N}_{\text{a}}$ are determined in analogy to Eq. (4).

[III]Since the residuals of the lines in $\mathscr{N}_{\text{p}}$ cannot be changed, it is sufficient to restrict $d_{\max}$ to those lines satisfying $B_j = \mathcal{B}$.

[IV]At this point, the values of the defects are the same as in stage **S.6.5.3**.

[V]For convenience, steps **S.6.5.8**–**S.6.5.11** will be referred to as the *exclusion procedure* and denoted with `ex_proc`($\mathscr{N}_{\text{a}}$, $\mathscr{N}_{\text{F}}$) in the remaining of this Appendix. During the exclusion procedure, lines are excluded *one by one* from $\mathscr{N}_{\text{a}}$ so as to avoid increasing the number of components in $\mathscr{N}_{\text{a}}$.

[VI]It is the number of energy levels in the jth component of $\mathscr{N}_{\text{a}}$.

Appendix C Stages of the local synchronization procedure (see also 2.4.3)

S.1 set flag $\mathcal{P}$ to 1;[I]

S.2 store the indices of the energy levels in $m_1, m_2, \ldots, m_{N_{\text{L}}}$ in the order that either $\text{comp}(m_i) < \text{comp}(m_j)$ or $\text{comp}(m_i) = \text{comp}(m_j)$ and $\bar{E}_{m_i} \le \bar{E}_{m_j}$ is valid for all $1 \le i \le j \le N_{\text{L}}$;

S.3 for all $1 \le i \le N_{\text{L}}$:

S.3.1 set $M = m_i$;

S.3.2 if $b_M = 100$, then **continue S.3**;

S.3.3 build a $\mathscr{N}_{\text{loc}}$ *local subnetwork*[II] of the SN in the following way: if $\text{up}(j) = M$, then set $P_j^{\langle \mathscr{N}_{\text{loc}} \rangle} = 1$, otherwise set $P_j^{\langle \mathscr{N}_{\text{loc}} \rangle} = 0$ for all $1 \le j \le N_{\text{T}}$;

S.3.4 save the indices of lines in $\mathscr{N}_{\rm loc}$ into $\mathscr{L}_1, \mathscr{L}_2, ..., \mathscr{L}_{N_{\rm loc}}$, where $N_{\rm loc}$ is the number of transitions in $\mathscr{N}_{\rm loc}$;

S.3.5 for all $1 \le j \le N_{\rm loc}$: set $\tilde{E}_j = \sigma_{\mathscr{L}_j} + \tilde{E}_{{\rm low}(\mathscr{L}_j)}$, which is the *upper trial energy* (UTE) of the Mth energy level from the $\mathscr{L}_j$th line;

S.3.6 initialize $n_{\rm expl}$ as 0;

S.3.7 for all $1 \le j \le N_{\rm T}$ with $B_j \le b_M$ and $M \in \{{\rm up}(j), {\rm low}(j)\}$: set $n_{\rm expl}$ to $n_{\rm expl}+1$;[III]

S.3.8 permute the UTEs such that $\tilde{E}_{p_j} \le \tilde{E}_{p_k}$ should be met for all $1 \le j \le k \le N_{\rm loc}$;

S.3.9 set $n_{\rm mrct} = \max(\max(n_{\rm expl}, \mu_{\rm sync}), \phi_{\rm td}{\rm TD}_M)$, where

(a) $n_{\rm mrct}$ is the *minimally required number of confirmative transitions*,[IV],

(b) $\mu_{\rm sync} \ge 2$ is the *synchronization margin*[V] ($\mu_{\rm sync} = 3$ is recommended)

(c) ${\rm TD}_M = |\{j \in N_{\rm T} : M \in \{{\rm up}(j), {\rm low}(j)\}\}|$ is the *transition degree* related to the Mth energy level in the SN, and

(d) $\phi_{\rm td}(=0.1)$ is the *transition degree factor*;

S.3.10 set $n_{\rm ct} = N_{\rm loc}$, where $n_{\rm ct}$ is the *number of confirmative* transitions;[VI]

S.3.11 for all $1 \le k \le N_{\rm loc} - n_{\rm mrct}$:

S.3.11.1 set the following parameters:

$$\begin{aligned} p_{\rm min} &= p_1, \\ \tilde{E}_{\rm min} &= \tilde{E}_{p_1}, \\ \tilde{E}_{\rm max} &= \tilde{E}_{p_{n_{\rm ct}}}, \\ \mathcal{D}_{\rm min} &= \tilde{E}_{p_{n_{\rm ct}-1}} - \tilde{E}_{\rm min}, \\ \mathcal{D}_{\rm max} &= \tilde{E}_{\rm max} - \tilde{E}_{p_2}, \\ \mathcal{D}_k &= \tilde{E}_{\rm max} - \tilde{E}_{\rm min}; \end{aligned} \tag{19}$$

S.3.11.2 if $\mathcal{D}_{\rm min} > \mathcal{D}_{\rm max}$, then

S.3.11.2.1 for all $1 \le l \le n_{\rm ct}-1$: set p_l to p_l+1;

S.3.11.2.2 replace $p_{n_{\rm ct}}$ with $p_{\rm min}$;

S.3.11.3 decrease $n_{\rm ct}$ to $n_{\rm ct}-1$;

S.3.12 set $n_{{\rm ct},0} = n_{\rm ct}$ and $\delta_{0,\max} = \max_{k=1}^{n_{\rm ct}} \delta_{0,\mathscr{L}_{p_k}}$;

S.3.13 for all $1 \le k \le N_{\rm loc} - n_{{\rm ct},0}$: if $\mathcal{D}_{N_{\rm loc}-n_{{\rm ct},0}-k+1} \le \phi_{\rm le}\delta_{0,\max}$, then increase $n_{\rm ct}$ to $n_{\rm ct}+1$, where $\phi_{\rm le}(=2)$ is the *local extension factor*;[VII]

S.3.14 if $n_{\rm ct} \ge n_{\rm mrct}$, then:

S.3.14.1 perform the following assignments:

$$\begin{aligned} \tilde{E}_{\rm min} &= \min_{k=1}^{n_{\rm ct}} \tilde{E}_{p_k}, \\ \tilde{E}_{\rm max} &= \min_{k=1}^{n_{\rm ct}} \tilde{E}_{p_k}, \\ \tilde{E}_{\rm mid} &= (\tilde{E}_{\rm min} + \tilde{E}_{\rm max})/2, \\ \varepsilon_{\rm mid} &= (\tilde{E}_{\rm max} - \tilde{E}_{\rm min})/2, \end{aligned} \tag{20}$$

where $\tilde{E}_{\rm mid}$ and $\varepsilon_{\rm mid}$ are the *midpoint energy* and its *uncertainty*, respectively;

S.3.14.2 if $|\bar{E}_M - \tilde{E}_{\rm mid}| > \max(\phi_{\rm sync}\varepsilon_{\rm mid}, \phi_{\rm ld})$, where $\phi_{\rm sync}(=5)$ denotes the *synchronization factor*, then[VIII] replace $\bar{E}_M$ with $\tilde{E}_{\rm mid}$ and reset $\mathcal{P}$ to 0;

S.4 if $\mathcal{P}=0$, then:

S.4.1 initialize $\mathscr{N}_{\rm F}$ as $P_j^{\langle\mathscr{N}_{\rm F}\rangle} = 0$ for all $1 \le j \le N_{\rm T}$;

S.4.2 for all $1 \le j \le N_{\rm T}$:

S.4.2.1 update Δ_j and d_j (see section 2.2);

S.4.2.2 if $P_j = 0$, then:

S.4.2.2.1 reset $\delta_j = \max(\delta_{0,j}, \phi_{\rm inc}|\Delta_j|)$;

S.4.2.2.2 if $\delta_j \le \phi_{\rm fil}{\rm RSU}^{[\mathfrak{s}(\mathfrak{i}_j)]}$, then set $P_j = 1$ and $B_j = \lfloor \log_{10}({\rm RSU}^{[\mathfrak{s}(\mathfrak{i}_j)]}) \rfloor$;[X]

S.4.2.3 if $d_j > 0$, then set $P_j^{\langle\mathscr{N}_{\rm F}\rangle} = 1$;[IX]

S.4.3 call `ex_proc`$(\mathscr{N}_{\rm LS}, \mathscr{N}_{\rm F})$;

S.5 **end** procedure.

Notes:

[I]Upon completion of the local synchronization procedure, $\mathcal{P}$ will designate whether all the constrained empirical energy levels are confirmed by this algorithm ($\mathcal{P}=1$) or there are energy values modified during the synchronization process ($\mathcal{P}=0$).

[II]The local subnetwork defined in stages **S.3.3**–**S.3.4** contains all the transitions of the *whole* SN which are included in the gsCD of the Mth energy level.

[III]The $n_{\rm expl}$ parameter represents the number of exploited lines involved in the Mth energy level. The jth transition of the SN is *exploited* if $B_j \le \max(b_{{\rm up}(j)}, b_{{\rm low}(j)})$, that is, if this line is utilized during the determination of its upper or lower energy level.

[IV]The $n_{\rm mrct}$ variable measures how many transitions should minimally be included in $\mathscr{N}_{\rm loc}$ to definitely confirm or reject the energies obtained from the $n_{\rm expl}$ exploited lines.

[V]$\mu_{\rm sync}$ provides a reasonable restriction to $n_{\rm mrct}$ in cases where $n_{\rm expl}$ has a very small value.

[VI]During execution of stages **S.3.10**–**S.3.11**, a collection of $n_{\rm ct} \ge n_{\rm mrct}$ transitions, called *confirmative linelist* (CL), is choosen from $\mathscr{N}_{\rm loc}$, such that (a) $n_{\rm ct}$ and the maximum unsigned deviation of the UTEs should be minimal in the CL, and (b) if possible, this largest unsigned UTE deviation must be greater than zero. In the first step, CL is filled with all the transitions of $\mathscr{N}_{\rm loc}$, then the 'worst' line, which is most distant from the remaining CL transitions in terms of its UTE, is left out *one by one* from the CL.

[VII]The CL list is extended with some further lines up to a well-defined limit for the UTE deviations.

[VIII]If $N_{\rm loc} \ge n_{\rm mrct}$ and $\bar{E}_M$ is at a larger distance from $\tilde{E}_{\rm mid}$ than a well-defined tolerance value, then $\tilde{E}_{\rm mid}$ is used instead of $\bar{E}_M$ in the rest of the local synchronization process.

[IX]If there are energy levels adjusted by the local synchronization procedure, then there must be lines of positive (updated) defects, for which the `ex_proc` algorithm needs to be called.

[X]In the case that certain constrained empirical energy levels were altered during the local synchronization, there may be transitions which can be reincluded in $\mathscr{N}_{\rm LS}$.

Appendix D Stages of the ABR procedure (see Section 2.4.4)

S.1 before the execution of the ABR method,[I] place the segments into the $R(1), R(2), ..., R(N_{\rm R})$ *recalibration classes*, where $N_{\rm R}$ is the number of recalibration classes, as follows:

(a) those segments which must be recalibrated with the same factor are grouped into the same class, and

(b) all the segments for which *simple recalibration* (recalibration with a single factor) is not permitted, are distributed into separate classes;

S.2 set $\chi = \mathcal{P}$;[II]

S.3 for all $1 \le i \le N_{\rm R}$: determine the ${\rm ESU}_{\rm min}^{[\![R(i)]\!]}$ values in the following way:

$${\rm ESU}_{\rm min}^{[\![R(i)]\!]} = \min_{\substack{j=1 \\ R(\mathfrak{R}_j)=R(i)}}^{N_{\mathfrak{s}}} {\rm ESU}^{[\mathfrak{s}(j)]}, \tag{21}$$

where $\mathfrak{R}_j$ is the *recalibration class index* of the $\mathfrak{s}(j)$ segment, for which $R(\mathfrak{R}_j)$ contains $\mathfrak{s}(j)$;

S.4 rearrange the recalibration classes such that ${\rm ESU}_{\rm min}^{[\![R(q_i)]\!]} \le {\rm ESU}_{\rm min}^{[\![R(q_j)]\!]}$

should be satisfied for all $1 \le i \le j \le N_R$;[III]

S.5 set the $\mathcal{F}$ flag[IV] to 1;

S.6 for $1 \le i \le N_R$:

S.6.1 set $\mathcal{R}$ to $R(q_i)$;

S.6.2 initialize $f^{[\![\mathcal{R}]\!]}$ as 1, where $f^{[\![\mathcal{R}]\!]}$ is the *recalibration factor* for $\mathcal{R}$;

S.6.3 if $\chi = 1$, then:[V]

S.6.3.1 construct a $\mathscr{N}_p$ subnetwork in the following form: $P_j^{\langle \mathscr{N}_p \rangle} = 0$ for all $1 \le j \le N_T$;

S.6.3.2 set $N_{p,c} = N_L$;

S.6.3.3 for all $1 \le j \le N_L$: assign $b_j = 100$, $\mathrm{core}_p(j) = \mathrm{comp}_p(j) = j$, and $\beta_j = 0$;

S.6.3.4 for all $B_{min} \le \mathcal{B} \le B_{max}$:

S.6.3.4.1 build a $\mathscr{N}_a$ subnetwork of $\mathscr{N}_{LS}$ according to the following scheme: if $B_j \le \mathcal{B}$, then $P_j^{\langle \mathscr{N}_a \rangle} = 1$, otherwise $P_j^{\langle \mathscr{N}_a \rangle} = 0$ for all $1 \le j \le N_T$;

S.6.3.4.2 if $\mathbf{P}^{\langle \mathscr{N}_a \rangle} = \mathbf{P}^{\langle \mathscr{N}_p \rangle}$, then **continue S.6.3.4**;

S.6.3.4.3 find the components of the $\mathscr{N}_a$ subnetwork;

S.6.3.4.4 save the $\mathrm{comp}_a(1), \mathrm{comp}_a(2), ..., \mathrm{comp}_a(N_L)$ component indices;

S.6.3.4.5 yield the solution for Eqs. (11)–(12);

S.6.3.4.6 calculate $\bar{\mathbf{E}}$ with the help of Eq. (9);

S.6.3.4.7 for all $1 \le j \le N_L$: if $b_j = 100$, $\mathscr{S}_a(\mathrm{comp}_a(j)) > 1$, and $\mathrm{core}_a(\mathrm{comp}_a(j)) = \mathrm{core}(\mathrm{comp}(j))$, then set $b_j = \mathcal{B}$;

S.6.3.4.8 overwrite $N_{p,c}$ with $N_{a,c}$;

S.6.3.4.9 for all $1 \le j \le N_{p,c}$: exchange $\mathrm{core}_p(j)$ with $\mathrm{core}_a(j)$;

S.6.3.4.10 for all $1 \le j \le N_L$: set $\mathrm{comp}_p(j) = \mathrm{comp}_a(j)$ and $\beta_j = \bar{E}_j$;

S.6.3.5 for all $1 \le j \le N_T$: update Δ_j;

S.6.3.6 set $\chi = 0$;

S.6.4 if simple recalibration is not permitted for $\mathcal{R}$, then **continue S.6**;[VI]

S.6.5 for all $1 \le j \le N_T$: if $R(\mathfrak{R}_{i_j}) = \mathcal{R}$ and $|\Delta_j| > \phi_{dec}\delta_j$, then set $P_j^{\langle \mathscr{N}_{reg} \rangle} = 1$, otherwise use $P_j^{\langle \mathscr{N}_{reg} \rangle} = 0$, where

(a) $\phi_{dec} (= 0.01)$ is the *declination factor*, and

(b) $\mathscr{N}_{reg}$ is a *regional subnetwork*[VII] of $\mathscr{N}_{LS}$;

S.6.6 set $n_{mrst} = \max(\lfloor \phi_{crdd} N_{reg} \rfloor, \mu_{rec})$,[VIII] where

(a) n_{mrst} is the *minimally required number of supportive transitions*,

(b) N_{reg} is the number of transitions in $\mathscr{N}_{reg}$,

(c) $\mu_{rec} \ge 2$ is the *recalibration margin* with the recommended value of $\mu_{rec} = 5$, and

(d) $\phi_{crdd} (= 0.5)$ is the *critical recalibration data-density factor*;

S.6.7 if $N_{reg} < n_{mrst}$, then **continue S.6**;[VI,IX]

S.6.8 save the indices of the lines in $\mathscr{N}_{reg}$ into $t_1, t_2, ..., t_{N_{reg}}$ such that $|\Delta_{t_j}| \le |\Delta_{t_k}|$ should be met for all $1 \le j \le k \le N_{reg}$;[X]

S.6.9 assign $n_{st} = n_{mrst}$, where n_{st} is the *number of supportive transitions*;

S.6.10 for all $n_{mrst} + 1 \le j \le N_{reg}$: if $|\Delta_{t_j}| \le \phi_{re}\mathrm{ESU}^{[\mathfrak{s}(\mathfrak{i}_{t_j})]}$, then increase n_{st} to $n_{st} + 1$, where $\phi_{re} (= 2)$ is the *regional extension factor*;

S.6.11 for all $1 \le j \le n_{st}$: specify the *pointwise recalibration factors* (PRF) as follows:[XI]

$$Pi_j = \frac{\bar{E}_{up(t_j)} - \bar{E}_{low(t_j)}}{\sigma_{t_j}}; \tag{22}$$

S.6.12 calculate the *median PRF value* as[XII]

$$\Pi_{med} = \underset{j=1}{\overset{n_{st}}{\mathrm{median}}}\, \Pi_j; \tag{23}$$

S.6.13 define the *recalibrated* and the *non-recalibrated average absolute resiuals* in the following fashion:[XIII]

$$\mathrm{AAR}_{rec}^{[\![\mathcal{R}]\!]} = \frac{1}{n_{st}} \left| \sum_{j=1}^{n_{st}} \left(\Pi_{med}\sigma_{t_j} - \bar{E}_{up(t_j)} + \bar{E}_{low(t_j)} \right) \right|, \tag{24}$$

$$\mathrm{AAR}_{nr}^{[\![\mathcal{R}]\!]} = \frac{1}{n_{st}} \left| \sum_{j=1}^{n_{st}} \Delta_{t_j} \right|; \tag{25}$$

S.6.14 if $\mathrm{AAR}_{nr}^{[\![\mathcal{R}]\!]} \le \max(\phi_{acc}\mathrm{ESU}_{min}^{[\![\mathcal{R}]\!]}, \phi_{ass}\mathrm{AAR}_{rec}^{[\![\mathcal{R}]\!]})$, then **continue S.6**,[VI,XIV] where

(a) $\phi_{acc} (= 0.8)$ is the *acceptability factor*, and

(b) $\phi_{ass} (= 3)$ is the *assimilation factor*;

S.6.15 set $\mathcal{F} = 0$;

S.6.16 assign $\chi = 1$;

S.6.17 reset $f^{[\![\mathcal{R}]\!]} = \Pi_{med}$;

S.6.18 for $1 \le j \le N_T$ with $R(\mathfrak{R}_{i_j}) = \mathcal{R}$:

S.6.18.1 replace σ_j with $f^{[\![\mathcal{R}]\!]}\sigma_j$;[XV]

S.6.18.2 update Δ_j;

S.6.19 for all $1 \le j \le N_{\mathfrak{s}}$ with $R(\mathfrak{R}_j) = \mathcal{R}$:

S.6.19.1 set $s = \mathfrak{s}(j)$;

S.6.19.2 save the indices of the non-excluded transitions in s into $u_1, u_2, ..., u_{N^{[s]}}$, where $N^{[s]}$ is the number of non-excluded lines in s;

S.6.19.3 provide the *recalibration RMSD* for the s segment as follows:

$$\mathrm{RMSD}_{rec}^{[s]} = \sqrt{\frac{1}{N^{[s]}} \sum_{k=1}^{N^{[s]}} \Delta_{u_k}^2}; \tag{26}$$

S.6.19.4 if $\mathrm{RMSD}_{rec}^{[s]} < \mathrm{RSU}^{[s]}$, then

S.6.19.4.1 substitute $\mathrm{RSU}^{[s]}$ with $\mathrm{RMSD}_{rec}^{[s]}$;[XVI];

S.6.19.4.2 for all $1 \le k \le N_T$ with $\mathfrak{s}(\mathfrak{i}_k) = s$:

S.6.19.4.2.1 if $\mathrm{RSU}^{[s]} = \mathrm{ESU}^{[s]}$, then set $\delta_{0,k} = \mathrm{RSU}^{[s]}$, else assign $\delta_{0,k} = \phi_{us}\mathrm{RSU}^{[s]}$;

S.6.19.4.2.2 set $B_k = \lfloor \log_{10}(\mathrm{RSU}^{[s]}) \rfloor$;

S.7 **end** procedure.

Notes:

[I]This partitioning of the segments can be performed right after the input files are processed.

[II]The χ boolean variable represents whether the $\bar{\mathbf{E}}$ vector should be corrected ($\chi = 1$) or not ($\chi = 0$) before a given recalibration class is recalibrated. If the entries of $\bar{\mathbf{E}}$ were changed during the local synchronization ($\mathcal{P} = 1$), they certainly have to be corrected.

[III]The recalibration classes are recalibrated in descending order of their minimum ESU values to reduce the distortion effects caused by highly uncertain spectral lines on the recalibration factors.

[IV]At stage **S.7**, $\mathcal{F}$ denotes whether $\mathscr{N}_{LS}$ is *fully recalibrated*, *i.e.*, none of the recalibration classes are recalibrated during the ABR procedure.

[V]The constrained energy levels are recalculated in a similar way as described in Appendix B, the only difference is that the wavenumber uncertainties are not refined here.

[VI]These empirical conditions, namely steps **S.6.4**, **S.6.7**, and **S.6.14**, have to be violated so that the $\mathcal{R}$ class can be recalibrated with the ABR algorithm.

[VII]$\mathscr{N}_{reg}$ contains all the transitions from the $\mathcal{R}$ class having unsigned residuals that are significantly larger than the empirical limit specified by the wavenumber uncertainties.

[VIII]It is the minimum number of transitions utilized during the determination of the $f^{[\![\mathcal{R}]\!]}$ factor.

[IX]If this condition is true, then there are too few lines in $\mathscr{N}_{reg}$ for the safe recalibration of $\mathcal{R}$.

[X]In steps **S.6.8**, **S.6.9**, and **S.6.10**, a set of $n_{st} \le n_{mrst}$ lines, called *supportive linelist (SL)*, is introduced, whose transitions are in best agreement in terms of their unsigned residuals.

[XI]The PRF values provide an interval, with the help of which the $f^{[\![\mathcal{R}]\!]}$ factor can be estimated.
[XII]It is found that the median PRF value is quite stable even against not too distant outliers.
[XIII]In an ideal case, both $\mathrm{AAR}_{\mathrm{nr}}^{[\![\mathcal{R}]\!]}$ and $\mathrm{AAR}_{\mathrm{rec}}^{[\![\mathcal{R}]\!]}$ should be close to zero.
[XIV]In the case that the $\mathcal{R}$ class is well calibrated, the condition of step **S.6.14** must be satisfied.
[XV]If $\mathcal{R}$ is recalibrated, then the wavenumbers of its lines should be multiplied with $f^{[\![\mathcal{R}]\!]}$.
[XVI]If necessary, the RSU values and the initial line uncertainties of the recalibrated $\mathcal{R}$ class should be modified to properly describe the accuracy of the improved wavenumbers.

Appendix E Stages of the FNC technique (see Section 2.4.5)

S.1 for all $1 \le i \le N_{\mathrm{T}}$: if $P_i = 0$ and $\delta_i \le \phi_{\mathrm{rei}}\mathrm{RSU}^{[\mathfrak{s}(\mathrm{i}_i)]}$, then reset $P_i = 1$, where $\phi_{\mathrm{rei}}(=100)$ is referred to as the *reinclusion factor*;[I]

S.2 detect the bridges of $\mathscr{N}_{\mathrm{LS}}$ *via* the BFS method;

S.3 for all $1 \le i \le N_{\mathrm{T}}$: if $P_i = 1$ and the ith transition is a bridge of $\mathscr{N}_{\mathrm{LS}}$, then set $\mathscr{B}_i = 1$, otherwise use $\mathscr{B}_i = 0$;

S.4 build the *maximum bridgeless subnetwork* (MBS)[II] of $\mathscr{N}_{\mathrm{LS}}$, designated with $\mathscr{N}_{\mathrm{B}}$, in the following way: $P_i^{\langle\mathscr{N}_{\mathrm{B}}\rangle} = P_i(1-\mathscr{B}_i)$ for all $1 \le i \le N_{\mathrm{T}}$;

S.5 identify the components of $\mathscr{N}_{\mathrm{B}}$ (*i.e.*, the *bridge components* of $\mathscr{N}_{\mathrm{LS}}$);

S.6 store the component indices of $\mathscr{N}_{\mathrm{B}}$ in $\mathrm{comp}_{\mathrm{B}}(1), \mathrm{comp}_{\mathrm{B}}(2), \ldots, \mathrm{comp}_{\mathrm{B}}(N_{\mathrm{L}})$;

S.7 save the component sizes of $\mathscr{N}_{\mathrm{B}}$ into $\mathscr{S}_{\mathrm{B}}(1), \mathscr{S}_{\mathrm{B}}(2), \ldots, \mathscr{S}_{\mathrm{B}}(N_{\mathrm{B,c}})$, where $N_{\mathrm{B,c}}$ is the number of components in $\mathscr{N}_{\mathrm{B}}$;

S.8 for all $1 \le i \le N_{\mathrm{L}}$:

S.8.1 specify the $\mathscr{R}_i$ *resistance*[III] of the ith energy level as

$$\mathscr{R}_i = \begin{cases} \text{'protected', if } \mathrm{comp}_{\mathrm{B}}(i) = \mathrm{comp}_{\mathrm{B}}(\mathrm{core}(\mathrm{comp}(i))), \\ \text{'unprotected', if } \mathscr{S}_{\mathrm{B}}(\mathrm{comp}_{\mathrm{B}}(i)) = 1, \\ \text{'semiprotected', otherwise;} \end{cases} \tag{27}$$

S.8.2 define the following 'connectivity' parameters,[IV]

$$\begin{aligned} \mathrm{LTD}_i &= \left|\left\{j : 1 \le j \le N_{\mathrm{T}}, P_j = 1, \text{ and } i \in \{\mathrm{up}(j), \mathrm{low}(j)\}\right\}\right|, \\ \mathrm{BTD}_i &= \left|\left\{j : 1 \le j \le N_{\mathrm{T}}, B_j \le \max(b_{\mathrm{up}(j)}, b_{\mathrm{low}(j)}), \text{ and } i \in \{\mathrm{up}(j), \mathrm{low}(j)\}\right\}\right|, \\ \mathrm{LSD}_i &= \left|\left\{\mathcal{S}(j) : 1 \le j \le N_{\mathrm{T}}, P_j = 1, \text{ and } i \in \{\mathrm{up}(j), \mathrm{low}(j)\}\right\}\right|, \\ \mathrm{BSD}_i &= \left|\left\{\mathcal{S}(j) : 1 \le j \le N_{\mathrm{T}}, B_j \le \max(b_{\mathrm{up}(j)}, b_{\mathrm{low}(j)}), \text{ and } i \in \{\mathrm{up}(j), \mathrm{low}(j)\}\right\}\right|, \end{aligned} \tag{28}$$

where the ith rovibrational energy level is associated with the LTD_i *leading transition degree*, the BTD_i *block transition degree*, the LSD_i *leading source degree*, and the BSD_i *block source degree*;

S.8.3 assign the *dependability grade*[V] to the ith energy level as follows:

$$\gamma_i = \begin{cases} \mathrm{A}^+, \text{ if } \mathscr{R}_i = \text{'protected'}, \mathrm{LSD}_i \ge \mathrm{LSD}_{\mathrm{crit}}, \text{ and } \mathrm{LTD}_i \ge \mathrm{LTD}_{\mathrm{crit}}, \\ \mathrm{A}^-, \text{ if } \mathscr{R}_i = \text{'protected'}, \mathrm{LSD}_i \ge \mathrm{LSD}_{\mathrm{crit}}, \text{ and } \mathrm{LTD}_i < \mathrm{LTD}_{\mathrm{crit}}, \\ \mathrm{B}^+, \text{ if } \mathscr{R}_i = \text{'protected'}, \mathrm{LSD}_i < \mathrm{LSD}_{\mathrm{crit}}, \text{ and } \mathrm{LTD}_i \ge \mathrm{LTD}_{\mathrm{crit}}, \\ \mathrm{B}^-, \text{ if } \mathscr{R}_i = \text{'protected'}, \mathrm{LSD}_i < \mathrm{LSD}_{\mathrm{crit}}, \text{ and } \mathrm{LTD}_i < \mathrm{LTD}_{\mathrm{crit}}, \\ \mathrm{C}^+, \text{ if } \mathscr{R}_i = \text{'semiprotected'}, \\ \mathrm{C}^-, \text{ if } \mathscr{R}_i = \text{'unprotected' and } b_i \ne 100, \\ \mathrm{D}, \text{ otherwise,} \end{cases} \tag{29}$$

where $\mathrm{LTD}_{\mathrm{crit}}(=10)$ and $\mathrm{LSD}_{\mathrm{crit}}(=6)$ are the *critical LTD and LSD*, respectively;

S.8.4 if $\mathrm{core}(\mathrm{comp}(i)) = i$, then set $\varepsilon_i = 0$[VI] and **continue S.8**;

S.8.5 construct a $\mathscr{N}_{\mathrm{frac}}$ *fractional subnetwork*[VII] of $\mathscr{N}_{\mathrm{LS}}$ in the following fashion: if $P_j = 1$, $\mathscr{B}_j = 0$, and $i \in \{\mathrm{up}(j), \mathrm{low}(j)\}$, then set $P_j^{\langle\mathscr{N}_{\mathrm{frac}}\rangle} = 1$, otherwise use $P_j^{\langle\mathscr{N}_{\mathrm{frac}}\rangle} = 0$ for all $1 \le j \le N_{\mathrm{T}}$;

S.8.6 save the indices of the transitions in $\mathscr{N}_{\mathrm{frac}}$ into $v_1, v_2, \ldots, v_{N_{\mathrm{frac}}}$ such that $\delta_{v_j} \le \delta_{v_k}$ holds for all $1 \le j \le k \le N_{\mathrm{frac}}$, where N_{frac} is the number of lines in $\mathscr{N}_{\mathrm{frac}}$;

S.8.7 if $N_{\mathrm{frac}} < n_{\mathrm{cwt}}$ or $|S^{(N_{\mathrm{frac}})}| < n_{\mathrm{cws}}$, then set $\varepsilon_i = -1$[VIII] and **continue S.8**, where

(a) $S^{(N_{\mathrm{frac}})} = \{\mathcal{S}(\mathcal{I}_{v_j}) : 1 \le j \le N_{\mathrm{frac}}\}$ is the set of sources including transitions with the v_1, v_2, ..., $v_{N_{\mathrm{frac}}}$ indices,
(b) $\mathcal{S}(l)$ is the lth source of the SN,
(c) $\mathcal{I}_m$ is the mth *source index* for which $\mathcal{S}(\mathcal{I}_m)$ contains the mth transition,
(d) $n_{\mathrm{cwt}}(=5)$ is the *critical number of witness transitions*, and
(e) $n_{\mathrm{cws}}(=3)$ is the *critical number of witness sources*;

S.8.8 introduce the *minimally required number of witness transitions*,[IX] n_{mrwt}, in this way:

$$n_{\mathrm{mrwt}} = \min\left\{k : n_{\mathrm{cwt}} \le k \le N_{\mathrm{frac}} \text{ and } |S^{(k)}| \ge n_{\mathrm{cws}}\right\}; \tag{30}$$

S.8.9 specify the *number of witness transitions*,[X] n_{wt}, in the following form:

$$n_{\mathrm{wt}} = \max\left\{k : n_{\mathrm{mrwt}} \le k \le N_{\mathrm{frac}} \text{ and } \delta_{v_k} \le \phi_{\mathrm{fe}}\delta_{v_{n_{\mathrm{mrwt}}}}\right\}, \tag{31}$$

where $\phi_{\mathrm{fe}}(=3)$ is the *fractional extension factor*;

S.8.10 estimate the ε_i uncertainty as[XI]

$$\varepsilon_i = \sqrt{\frac{1}{n_{\mathrm{wt}}}\sum_{k=1}^{n_{\mathrm{wt}}}\delta_{v_k}^2}; \tag{32}$$

S.9 for all $1 \le i \le N_{\mathfrak{s}}$:

S.9.1 set $s = \mathfrak{s}(i)$;

S.9.2 introduce the following statistical quantities:[XII]

$$\begin{aligned} \sigma_{\min}^{[s]} &= \min_{\substack{j=1 \\ \mathfrak{s}(\mathrm{i}_j)=s}}^{N_{\mathrm{T}}} \sigma_j, \\ \sigma_{\max}^{[s]} &= \max_{\substack{j=1 \\ \mathfrak{s}(\mathrm{i}_j)=s}}^{N_{\mathrm{T}}} \sigma_j, \\ \mathrm{ASU}^{[s]} &= \frac{1}{V^{[s]}}\sum_{\substack{j=1 \\ \mathfrak{s}(\mathrm{i}_j)=s}}^{N_{\mathrm{T}}} P_j\delta_j, \\ \mathrm{MSU}^{[s]} &= \max_{\substack{j=1 \\ \mathfrak{s}(\mathrm{i}_j)=s \\ P_j=1}}^{N_{\mathrm{T}}} \delta_j, \end{aligned} \tag{33}$$

where $\mathrm{ASU}^{[s]}$ and $\mathrm{MSU}^{[s]}$ are the *average and maximum segment uncertainties* of the s segment, respectively;

S.9.3 provide the $(A^{[s]}, V^{[s]}, E^{[s]})$ triplet[XIII] for the s segment as

$$\begin{aligned} A^{[s]} &= \left|\left\{j : 1 \le j \le N_{\mathrm{T}} \text{ and } \mathfrak{s}(\mathrm{i}_j) = s\right\}\right|, \\ V^{[s]} &= \left|\left\{j : 1 \le j \le N_{\mathrm{T}}, P_j = 1, \text{ and } \mathfrak{s}(\mathrm{i}_j) = s\right\}\right|, \\ E^{[s]} &= \left|\left\{j : 1 \le j \le N_{\mathrm{T}}, P_j = 1, B_j \le \max(b_{\mathrm{up}(j)}, b_{\mathrm{low}(j)}), \text{ and } \mathfrak{s}(\mathrm{i}_j) = s\right\}\right|, \end{aligned} \tag{34}$$

where $A^{[s]}$, $V^{[s]}$, and $E^{[s]}$ symbolize the *number of assigned, validated, and exploited transitions* corresponding to the s segment, respectively;

S.10 for all $1 \le i \le N_{\mathrm{T}}$: represent the *dependability grade* of the ith transition, Γ_i, with the lower[V] grade of $\gamma_{\mathrm{up}(i)}$ and $\gamma_{\mathrm{low}(i)}$;

S.11 **end** procedure.

Notes:

[I]Although the reincluded lines may have relatively large uncertainties, they could be useful for corroborating the dependability of the energy levels.

[II]$\mathscr{N}_{\mathrm{B}}$ is built by placing the non-bridge transitions of $\mathscr{N}_{\mathrm{LS}}$ into this subnetwork.

[III]The resistance of the ith rovibrational energy level is (a) *protected*, if it is included in the same bridge component as the core of its component in $\mathscr{N}_{\mathrm{LS}}$, (b) *unprotected*, if it is alone in its bridge component, or (c) *semiprotected*, otherwise.

[IV]These quantities can be interpreted in the following way: (a) LTD_i is the total number of lines in $\mathscr{N}_{\mathrm{LS}}$ incident to the ith energy level, (b) BTD_i is the number of exploited lines from $\mathscr{N}_{\mathrm{LS}}$ including this rovibrational state, and (c) LSD_i and BSD_i correspond to the numbers of sources containing transitions used in the definitions of LTD_i and BTD_i, respectively.

[V]The particular grades are listed starting from the best grade down to the worst one. In fact, energy levels with grade D are undefined, as there is no line in $\mathscr{N}_{\mathrm{LS}}$ incident to these energy levels.

[VI]The uncertainties of the cores are set to zero.

[VII]$\mathscr{N}_{\mathrm{frac}}$ contains all the non-bridge lines from $\mathscr{N}_{\mathrm{LS}}$ which are incident to the ith energy level.

[VIII]$\varepsilon_i = -1$ indicates that a reliable uncertainty cannot be assigned to the ith energy level.

[IX]$n_{\mathrm{mrwt}} \geq n_{\mathrm{cwt}}$ is the smallest positive integer for which transitions indexed with $\nu_1, \nu_2, \ldots, \nu_{n_{\mathrm{mrwt}}}$ arise from at least n_{cws} sources.

[X]As a result of this stage, a collection of $n_{\mathrm{wt}} \geq n_{\mathrm{mrwt}}$ transitions, called *witness linelist* (WL), is obtained, whose lines are suitable for the estimation of the uncertainty of the ith energy level.

[XI]If $n_{\mathrm{mrwt}} > 0$ and the ith energy level is not a core, then ε_i is calculated *via* Eq. (32).

[XII]The $[\sigma_{\min}^{[s]}, \sigma_{\max}^{[s]}]$ interval represents the measurement range of the wavenumbers in the s segment, while the parameters $\mathrm{ASU}^{[s]}$ and $\mathrm{MSU}^{[s]}$ describe the estimated accuracy of s at the end of the intMARVEL procedure.

[XIII]This triplet can be denoted with 'A/V/E', as well.

Acknowledgements

The authors thank Dr. Csaba Fábri for fruitful discussions. The Budapest group gratefully acknowledges the financial support they received from NKFIH (grant number K119658). The Budapest group also received support from the grant VEKOP-2.3.2-16-2017-00014, supported by the European Union and the State of Hungary and co-financed by the European Regional Development Fund. The collaboration between ELTE and UCL was supported by the CM1405 COST action, MOLIM: Molecules in Motion.

References

1 J. L. Hall, Nobel Lecture: Defining and measuring optical frequencies, Rev. Mod. Phys. 78 (2006) 1279–1295.
2 T. W. Hänsch, Nobel Lecture: Passion for precision, Rev. Mod. Phys. 78 (2006) 1297–1309.
3 M. J. Thorpe, D. Balslev-Clausen, M. S. Kirchner, J. Ye, Cavity-enhanced optical frequency comb spectroscopy: application to human breath analysis, Opt. Express 16 (2008) 2387–2397.
4 A. Foltynowicz, P. Maslowski, A. J. Fleisher, B. J. Bork, J. Ye, Cavity-enhanced optical frequency comb spectroscopy in the mid-infrared application to trace detection of hydrogen peroxide, Appl. Phys. B 110 (2013) 163–175.
5 C. A. Alrahman, A. Khodabakhsh, F. M. Schmidt, Z. Qu, A. Foltynowicz, Cavity-enhanced optical frequency comb spectroscopy of high-temperature H_2O in a flame, Opt. Express 22 (2014) 13889–13895.
6 D. Z. Kandula, C. Gohle, T. J. Pinkert, W. Ubachs, S. E. Eikema, Extreme ultraviolet frequency comb metrology, Phys. Rev. Lett. 105 (2010) 063001.
7 D. Z. Kandula, C. Gohle, T. J. Pinkert, W. Ubachs, S. E. Eikema, XUV frequency comb metrology on the ground state of helium, Phys. Rev. A 84 (2011) 062512.
8 A. G. Császár, G. Czakó, T. Furtenbacher, E. Mátyus, An active database approach to complete spectra of small molecules, Ann. Rep. Comp. Chem. 3 (2007) 155–176.
9 A. G. Császár, T. Furtenbacher, Spectroscopic networks, J. Mol. Spectrosc. 266 (2011) 99–103.
10 A. G. Császár, T. Furtenbacher, P. Árendás, Small molecules – Big data, J. Phys. Chem. A 120 (2016) 8949–8969.
11 T. Furtenbacher, A. G. Császár, J. Tennyson, MARVEL: measured active rotational-vibrational energy levels, J. Mol. Spectrosc. 245 (2007) 115–125.
12 T. Furtenbacher, A. G. Császár, MARVEL: measured active rotational-vibrational energy levels. II. Algorithmic improvements, J. Quant. Spectrosc. Radiat. Transfer 113 (2012) 929–935.
13 R. Tóbiás, T. Furtenbacher, A. G. Császár, Cycle bases to the rescue, J. Quant. Spectrosc. Radiat. Transfer 203 (2017) 557–564.
14 T. Furtenbacher, I. Szabó, A. G. Császár, P. F. Bernath, S. N. Yurchenko, J. Tennyson, Experimental Energy Levels and Partition Function of the $^{12}C_2$ Molecule, Astrophys. J., Suppl. Ser. 224 (2016) 44.
15 J. Tennyson, P. F. Bernath, L. R. Brown, A. Campargue, M. R. Carleer, A. G. Császár, R. R. Gamache, J. T. Hodges, A. Jenouvrier, O. V. Naumenko, O. L. Polyansky, L. S. Rothman, R. A. Toth, A. C. Vandaele, N. F. Zobov, L. Daumont, A. Z. Fazliev, T. Furtenbacher, I. F. Gordon, S. N. Mikhailenko, S. V. Shirin, Critical evaluation of the rotational-vibrational spectra of water vapor. Part I. Energy levels and transition wavenumbers for $H_2{}^{17}O$ and $H_2{}^{18}O$, J. Quant. Spectrosc. Radiat. Transfer 110 (2009) 573–596.
16 J. Tennyson, P. F. Bernath, L. R. Brown, A. Campargue, M. R. Carleer, A. G. Császár, R. R. Gamache, J. T. Hodges, A. Jenouvrier, O. V. Naumenko, O. L. Polyansky, L. S. Rothman, R. A. Toth, A. C. Vandaele, N. F. Zobov, A. Z. Fazliev, T. Furtenbacher, I. F. Gordon, S.-M. Hu, S. N. Mikhailenko, B. Voronin, Critical evaluation of the rotational-vibrational spectra of water vapor. Part II. Energy levels and transition wavenumbers for $HD^{16}O$, $HD^{17}O$, and $HD^{18}O$, J. Quant. Spectrosc. Radiat. Transfer 110 (2010) 2160–2184.
17 J. Tennyson, P. F. Bernath, L. R. Brown, A. Campargue, A. G. Császár, L. Daumont, R. R. Gamache, J. T. Hodges, O. V. Naumenko, O. L. Polyansky, L. S. Rothman, A. C. Vandaele, N. F. Zobov, A. R. Al Derzi, C. Fábri, A. Z. Fazliev, T. Furtenbacher, I. E. Gordon, L. Lodi, I. I. Mizus, IUPAC critical evaluation of the rotational-vibrational spectra of water vapor. Part III: Energy levels and transition wavenumbers for $H_2{}^{16}O$, J. Quant. Spectrosc. Rad. Transfer 117 (2013) 29–58.
18 J. Tennyson, P. F. Bernath, L. R. Brown, A. Campargue, A. G. Császár, L. Daumont, R. R. Gamache, J. T. Hodges, O. V. Naumenko, O. L. Polyansky, L. S. Rothmam, A. C. Vandaele, N. F. Zobov, N. Dénes, A. Z. Fazliev, T. Furtenbacher, I. E. Gordon, S.-M. Hu, T. Szidarovszky, I. A. Vasilenko, IUPAC critical evaluation of the rotational-vibrational spectra of water vapor. Part IV. Energy levels and transition wavenumbers for $D_2{}^{16}O$, $D_2{}^{17}O$ and $D_2{}^{18}O$, J. Quant. Spectrosc. Radiat. Transfer 142 (2014) 93–108.
19 J. Tennyson, P. F. Bernath, L. R. Brown, A. Campargue, A. G. Császár, L. Daumont, R. R. Gamache, J. T. Hodges, O. V. Naumenko, O. L. Polyansky, L. S. Rothman, A. C. Vandaele, N. F. Zobov, A Database of Water Transitions from Experiment and Theory (IUPAC Technical Report), Pure Appl. Chem. 86 (2014) 71–83, doi:10.1515/pac-2014-5012.
20 R. Tóbiás, T. Furtenbacher, A. G. Császár, O. V. Naumenko, J. Tennyson, J.-M. Flaud, P. Kumar, B. Poirier, Critical evaluation of measured rotational–vibrational transitions of four sulphur isotopologues of $S^{16}O_2$, J. Quant. Spectrosc. Radiat. Transfer 208 (2018) 152–163.
21 T. Furtenbacher, T. Szidarovszky, E. Mátyus, C. Fábri, A. G. Császár, Analysis of the Rotational-Vibrational States of the Molecular Ion H_3^+, J. Chem. Theory Comput. 9 (2013) 5471–5478.
22 T. Furtenbacher, T. Szidarovszky, C. Fábri, A. G. Császár, MARVEL Analysis of the Rotational-Vibrational States of the Molecular Ions H_2D^+ and D_2H^+, Phys. Chem. Chem. Phys. 15 (2013) 10181–10193.
23 A. R. Al Derzi, T. Furtenbacher, J. Tennyson, S. N. Yurchenko, A. G. Császár, MARVEL analysis of the measured high-resolution spectra of $^{14}NH_3$, J. Quant. Spectrosc. Rad. Transfer 161 (2015) 117 – 130.
24 C. Fábri, E. Mátyus, T. Furtenbacher, L. Nemes, B. Mihály, T. Zoltáni, A. G. Császár, Variational Quantum Mechanical and Active Database Approaches to the Rotational-Vibrational Spectroscopy of Ketene, J. Chem. Phys. 135 (2011) 094307.
25 C. Sousa-Silva, L. K. McKemmish, K. L. Chubb, J. Baker, E. J. Barton, M. N. Gorman, T. Rivlin, J. Tennyson, Original Research By Young Twinkle Students (ORBYTS): When can students start performing original research?, Phys. Educ. 53 (2018) 015020.
26 L. K. McKemmish, T. Masseron, S. Sheppard, E. Sandeman, Z. Schofield, T. Furtenbacher, A. G. Császár, J. Tennyson, C. Sousa-Silva, MARVEL analysis of the measured high-resolution spectra of $^{48}Ti^{16}O$, Astrophys. J., Suppl. Ser. 228 (2017) 15.
27 L. K. McKemmish, K. Goodhew, S. Sheppard, A. Bennet, A. Martin, A. Singh, C. Sturgeon, R. Godden, T. Furtenbacher, A. G. Császár, J. Tennyson, MARVEL analysis of the measured high-resolution spec-

tra of $^{90}Zr^{16}O$, Astrophys. J., Suppl. Ser. .
28 K. L. Chubb, O. V. Naumenko, S. Keely, S. Bartolotto, S. MacDonald, M. Mukhtar, A. Grachov, J. White, E. Coleman, A. Liu, A. Z. Fazliev, E. R. Polovtseva, V. M. Horneman, A. Campargue, T. Furtenbacher, A. G. Császár, S. N. Yurchenko, J. Tennyson, MARVEL analysis of the measured high-resolution rovibrational spectra of H_2S, J. Quant. Spectrosc. Radiat. Transfer 218 (2018) 178–186.
29 K. L. Chubb, M. Joseph, J. Franklin, N. Choudhury, T. Furtenbacher, A. G. Császár, G. Gaspard, P. Oguoko, A. Kelly, S. N. Yurchenko, J. Tennyson, C. Sousa-Silva, MARVEL analysis of the measured high-resolution spectra of C_2H_2, J. Quant. Spectrosc. Radiat. Transfer 204 (2018) 42–55.
30 I. Gordon, L. Rothman, C. Hill, R. V. Kochanov, Y. Tan, P. Bernath, M. Birk, V. Boudon, A. Campargue, K. V. Chance, B. Drouin, J.-M. Flaud, D. Gamache, R.R. Jacquemart, V. I. Perevalov, A. Perrin, M. A. H. Smith, J. Tennyson, H. Tran, V. G. Tyuterev, G. C. Toon, J. T. Hodges, K. P. Shine, A. Barbe, A. G. Császár, M. V. Devi, T. Furtenbacher, J. J. Harrison, A. Jolly, T. Johnson, T. Karman, I. Kleiner, A. Kyuberis, J. Loos, O. M. Lyulin, S. N. Mikhailenko, N. Moazzen-Ahmadi, H. S. P. Müller, O. V. Naumenko, A. V. Nikitin, O. L. Polyansky, M. Rey, M. Rotger, S. Sharpe, E. Starikova, S. A. Tashkun, J. Vander Auwera, G. Wagner, J. Wilzewski, P. Wcislo, S. Yu, E. Zak, The HITRAN2016 molecular spectroscopic database, J. Quant. Spectrosc. Radiat. Transfer 203 (2017) 3–69.
31 R. R. Gamache, C. Roller, E. Lopes, I. E. Gordon, L. S. Rothman, O. L. Polyansky, N. F. Zobov, A. A. Kyuberis, J. Tennyson, S. N. Yurchenko, A. G. Császár, T. Furtenbacher, X. Huang, D. W. Schwenke, T. J. Lee, B. J. Drouin, S. A. Tashkun, V. I. Perevalov, R. V. Kochanov, Total Internal Partition Sums for 167 isotopologues of 53 molecules important in planetary atmospheres: application to HITRAN2016 and beyond, J. Quant. Spectrosc. Radiat. Transfer 203 (2017) 70–87.
32 J. Tennyson, S. N. Yurchenko, ExoMol: molecular line lists for exoplanet and other atmospheres, Mon. Not. Royal Astron. Soc. 425 (2012) 21–33.
33 O. L. Polyansky, A. A. Kyuberis, L. Lodi, J. Tennyson, R. I. Ovsyannikov, N. Zobov, ExoMol molecular line lists XIX: high accuracy computed line lists for $H_2{}^{17}O$ and $H_2{}^{18}O$, Mon. Not. Royal Astron. Soc. 466 (2017) 1363–1371.
34 O. L. Polyansky, A. A. Kyuberis, N. F. Zobov, J. Tennyson, S. N. Yurchenko, L. Lodi, ExoMol molecular line lists XXX: a complete high-accuracy line list for water, Mon. Not. Royal Astron. Soc. doi: 10.1093/mnras/sty1877.
35 L. Lodi, J. Tennyson, Line lists for $H_2{}^{18}O$ and $H_2{}^{17}O$ based on empirically-adjusted line positions and ab initio intensities, J. Quant. Spectrosc. Radiat. Transfer 113 (2012) 850–858.
36 M. Birk, G. Wagner, J. Loos, L. Lodi, O. L. Polyansky, A. A. Kyuberis, N. F. Zobov, J. Tennyson, Accurate line intensities for water transitions in the infrared: comparison of theory and experiment, J. Quant. Spectrosc. Radiat. Transfer 203 (2017) 88–102.
37 A. A. Kyuberis, N. F. Zobov, O. V. Naumenko, B. A. Voronin, O. L. Polyansky, L. Lodi, A. Liu, S.-M. Hu, J. Tennyson, Room temperature linelists for deuterated water, J. Quant. Spectrosc. Radiat. Transfer 203 (2017) 175–185.
38 T. Furtenbacher, T. Szidarovszky, J. Hrubý, A. A. Kyuberis, N. F. Zobov, O. L. Polyansky, J. Tennyson, A. G. Császár, Definitive Ideal-Gas Thermochemical Functions of the $H_2{}^{16}O$ Molecule, J. Phys. Chem. Ref. Data 45 (2016) 043104.
39 I. Simkó, T. Furtenbacher, N. Dénes, T. Szidarovszky, J. Hrubý, N. F. Zobov, O. L. Polyansky, J. Tennyson, A. G. Császár, Recommended ideal-gas thermochemical functions for heavy water and its substituent isotopologues, J. Phys. Chem. Ref. Data 023104 (2017) 46.
40 C. Burgess, J. Hammond, Wavelength standards for the near-infrared spectral region, Spectroscopy 22 (2007) 40–48.
41 T. Furtenbacher, A. G. Császár, On employing $H_2{}^{16}O$, $H_2{}^{17}O$, $H_2{}^{18}O$, and $D_2{}^{16}O$ lines as frequency standards in the 15 – 170 cm^{-1} window, J. Quant. Spectrosc. Radiat. Transfer 109 (2008) 1234–1251.
42 H. M. Pickett, R. L. Poynter, E. A. Cohen, M. L. Delitsky, J. C. Pearson, H. S. P. Müller, Submillimeter, millimeter, and microwave spectral line catalog, J. Quant. Spectrosc. Radiat. Transfer 60 (1998) 883–890.
43 JPL catalog, URL = `http://spec.jpl.nasa.gov/ftp/pub/catalog/catdir.html`, 2018.
44 R. Lanquetin, L. H. Coudert, C. Camy-Peyret, High-lying rotational levels of water: An analysis of the energy levels of the five first vibrational states, J. Mol. Spectrosc. 206 (2001) 54–67.
45 D. Mondelain, S. N. Mikhailenko, E. V. Karlovets, S. Béguier, S. Kassi, A. Campargue, Comb-assisted cavity ring down spectroscopy of ^{17}O-enriched water between 7443 and 7921 cm^{-1}, J. Quant. Spectrosc. Radiat. Transfer 203 (2017) 206–212.
46 J. Chen, T.-P. Hua, L.-G. Tao, Y. Sun, A.-W. Liu, S.-M. Hu, Absolute frequencies of water lines near 790 nm with 10^{-11} accuracy, J. Quant. Spectrosc. Radiat. Transfer 205 (2018) 91–95.
47 S. G. Kukolich, Measurement of the molecular g values in H_2O and D_2O and hyperfine structure in H_2O, J. Chem. Phys. 50 (1969) 3751–3755.
48 G. Steenbeckeliers, J. Bellet, Spectre micro-onde de molecules $H_2{}^{16}O$, $H_2{}^{17}O$ et $H_2{}^{18}O$, C. R. Acad. Sc. Paris 273 (1971) 471–474.
49 C. K. Jen, Rotational Magnetic Moments in Polyatomic Molecules, Phys. Rev. 81 (1951) 197–203.
50 D. W. Posener, M. W. P. Strandberg, Centrifugal distortion effect in asymmetric top molecules III. H_2O, D_2O, and HDO, Phys. Rev. 95 (1954) 374–384.
51 J.-M. Flaud, C. Camy-Peyret, A. Valentin, Spectre infrarouge a haute résolution des bandes $\nu_1+\nu_2$ et $\nu_2+\nu_3$ de $H_2{}^{16}O$, J. Phys. 33 (1972) 741–747.
52 C. Huiszoon, A high resolution spectrometer for the shorter millimeter wavelength region, Rev. Sci. Instr. 42 (1971) 477–481.
53 G. Y. Golubiatnikov, V. N. Markov, A. Guarnieri, R. Knöchel, Hyperfine structure of $H_2{}^{16}O$ and $H_2{}^{18}O$ measured by Lamb-dip technique in the 180–560 GHz frequency range, J. Mol. Spectrosc. 240 (2006) 251–254.
54 F. C. De Lucia, P. Helminger, R. L. Cook, W. Gordy, Submillimeter microwave spectrum of $H_2{}^{16}O$, Phys. Rev. A 5 (1972) 487–490.
55 W. C. King, W. Gordy, One-to-two millimeter wave spectroscopy. IV. Experimental methods and results for OCS, CH_3F, and H_2O, Phys. Rev. 93 (1954) 407–412.
56 M. A. Koshelev, M. Y. Tretyakov, G. Y. Golubiatnikov, V. V. Parshin, V. N. Markov, I. A. Koval, Broadening and shifting of the 321-, 325- and 380-GHz lines of water vapor by pressure of atmospheric gases, J. Mol. Spectrosc. 241 (2007) 101–108.
57 G. Cazzoli, C. Puzzarini, M. E. Harding, J. Gauss, The hyperfine structure in the rotational spectrum of water: Lamb-dip technique and quantum-chemical calculations, Chem. Phys. Lett. 473 (2009) 21–25.
58 V. N. Markov, A. F. Krupnov, Measurements of the Pressure Shift of the (110)–(101) Water Line at 556 GHz Produced by Mixtures of Gases, J. Mol. Spectrosc. 172 (1995) 211–214.
59 F. Matsushima, H. Odashima, T. Iwasaki, S. Tsunekawa, K. Takagi, Frequency measurement of pure rotational transitions of H_2O from 0.5 to 5 THz, J. Mol. Spectrosc. 352/353 (1995) 371–378.
60 D. A. Stephenson, R. G. Strauch, Water vapor spectrum near 600 GHz, J. Mol. Spectrosc. 35 (1970) 494–495.
61 A. V. Burenin, T. M. Fevralskikh, E. N. Karyakin, O. L. Polyansky, S. M. Shapin, Effective Pade Hamiltonian operator and its application for treatment of $H_2{}^{16}O$ rotational spectrum in the ground state, J. Mol. Spectrosc. 100 (1983) 182–192.
62 S. Yu, J. C. Pearson, B. J. Drouin, Terahertz spectroscopy of water in its second triad, J. Mol. Spectrosc. 288 (2013) 7–10.
63 M. A. Koshelev, Collisional broadening and shifting of the (211)–(202) transition of $H_2{}^{16}O$, $H_2{}^{17}O$, $H_2{}^{18}O$ by atmosphere gases, J. Quant. Spectrosc. Rad. Transfer 112 (2011) 550–552.
64 P. Helminger, J. K. Messer, F. C. De Lucia, Continuously tunable coherent spectroscopy for the 0.1–1.0 THz region, Appl. Phys. Lett. 42 (1983) 309–10.
65 J. K. Messer, F. C. De Lucia, P. Helminger, The pure rotational spectrum of water vapor—A millimeter, submillimeter, and far infrared analysis, Int. J. Infr. Millim. Waves 4 (1983) 505–539.
66 A. Miani, J. Tennyson, Can ortho-para transitions for water be observed?, J. Chem. Phys. 120 (2004) 2732–2739.
67 I. Snellen, High-dispersion spectroscopy of extrasolar planets: from CO in hot Jupiters to O_2 in exo-Earths, Phil. Trans. Royal Soc. London A 372 (2014) 20130075.
68 H. J. Hoeijmakers, R. J. de Kok, I. A. G. Snellen, M. Brogi, J. L. Birkby, H. Schwarz, A search for TiO in the optical high-resolution transmission spectrum of HD 209458b: Hindrance due to inaccuracies in the line database, Astron. Astrophys. 575 (2015) A20.
69 J. L. Birkby, R. J. de Kok, M. Brogi, E. J. W. de Mooij, H. Schwarz, S. Albrecht, I. A. G. Snellen, Detection of water absorption in the day side atmosphere of HD 189733 b using ground-based high-resolution spectroscopy at 3.2 mu m, Mon. Not. Royal Astron. Soc. 436 (2013) L35–L39.
70 J. L. Birkby, R. J. de Kok, M. Brogi, H. Schwarz, I. A. G. Snellen, Discovery of water at high spectral resolution in the atmosphere of 51 Peg b, Astrophys. J. 153 (2017) 138.
71 J. Birkby, private communication, 2018.
72 https://www.craf.eu/iau-list-of-important-spectral-lines/#IAU%20list, 2018.
73 M. Herman, J. W. C. Johns, A. R. W. McKellar, High resolution laser Stark and infrared–radiofrequency double resonance spectroscopy of $H_2{}^{16}O$ at 6 μm, Can. J. Phys. 57 (1979) 397–401.

74 H. Kuze, Microwave spectrum of water in the v_2 excited vibrational state, Astrophys. J. 239 (1980) 1131–1133.
75 L. R. Brown, R. A. Toth, Comparison of the frequencies of NH_3, CO_2, H_2O, N_2O, CO, and CH_4 as infrared calibration standards, J. Opt. Soc. Am. B 2 (1985) 842–856.
76 O. I. Baskakov, V. A. Alekseev, E. A. Alekseev, B. I. Polevoi, New submillimeter rotational lines of water and its isotopes, Opt. i Spektrosk. 63 (5) (1987) 1016–1018.
77 S. P. Belov, I. N. Kozin, O. L. Polyansky, M. Y. Tretyakov, N. F. Zobov, Rotational spectrum of the $H_2{}^{16}O$ molecule in the (010) excited vibrational state, J. Mol. Spectrosc. 126 (1987) 113–117.
78 T. Amano, F. Scappini, Millimeter-wave spectrum of rotationally excited H_2O, Chem. Phys. Lett. 182 (1991) 93–95.
79 J. C. Pearson, T. Anderson, E. Herbst, F. C. De Lucia, P. Helminger, Millimeter- and submillimeter-wave spectrum of highly excited states of water, Astrophys. J. 379 (1991) L41–L43.
80 R. A. Toth, v_2 band of $H_2{}^{16}O$ - line strengths and transition frequencies, J. Opt. Soc. Am. B 8 (1991) 2236–2255.
81 R. A. Toth, $2v_2$–v_2 and $2v_2$ bands of $H_2{}^{16}O$, $H_2{}^{17}O$, and $H_2{}^{18}O$: line positions and strengths, J. Opt. Soc. Am. B 10 (1993) 1526–1544.
82 R. A. Toth, v_1–v_2, v_3–v_2, v_1 and v_3 bands of $H_2{}^{16}O$: line positions and strengths, J. Opt. Soc. Am. B 10 (1993) 2006–2029.
83 R. Paso, V.-M. Horneman, High-resolution rotational absorption spectra of $H_2{}^{16}O$, $HD^{16}O$, and $D_2{}^{16}O$ between 110 and 500 cm^{-1}, J. Opt. Soc. Am. B 12 (1995) 1813–1838.
84 J. C. Pearson, Ph.D. thesis Data from 00ChPePiMa.
85 S. P. Belov, private communication. Data from 00ChPePiMa.
86 L. R. Brown, J. S. Margolis, Empirical line parameters of NH_3 from 4791 to 5294 cm^{-1}, J. Quant. Spectrosc. Radiat. Transfer 56 (1996) 283–294.
87 P. D. Natale, L. Lorini, M. Inguscio, I. G. Nolt, J. H. Park, G. D. Lonardo, L. Fusina, P. A. R. Ade, A. G. Murray, Accurate frequency measurement for H_2O and $^{16}O_3$ in the 119 cm^{-1} OH atmospheric window, Appl. Opt. 36 (1997) 8526–8532.
88 S. N. Mikhailenko, V. G. Tyuterev, K. A. Keppler, B. P. Winnewisser, M. Winnewisser, G. Mellau, S. Klee, K. N. Rao, The $2v_2$ band of water: analysis of new FTS measurements and high-K_a transitions and energy levels, J. Mol. Spectrosc. 184 (1997) 330–349.
89 P. Chen, J. C. Pearson, H. M. Pickett, S. Matsuura, G. A. Blake, Submillimeter-wave measurements and analysis of the ground and $v_2 = 1$ states of water, Astrophys. J. Suppl. Ser. 128 (2000) 371–385.
90 Q. Zou, P. Varanasi, Laboratory measurement of the spectroscopic line parameters of water vapor in the 610–2100 and 3000–4050 cm^{-1} regions at lower-tropospheric temperatures, J. Quant. Spectrosc. Radiat. Transfer 82 (2003) 45–98.
91 V.-M. Horneman, R. Anttila, S. Alanko, J. Pietila, Transferring calibration from CO_2 laser lines to far infrared water lines with the aid of the v_2 band of OCS and the v_2, $v_1 - v_2$, and $v_1 + v_2$ bands of $^{13}CS_2$: molecular constants of $^{13}CS_2$, J. Mol. Spectrosc. 234 (2005) 238–254.
92 R. A. Toth, Measurements of positions, strengths and self-broadened widths of H_2O from 2900 to 8000 cm^{-1}: line strength analysis of the 2nd triad bands, J. Quant. Spectrosc. Radiat. Transfer 94 (2005) 51–107.
93 L. Joly, B. Parvitte, V. Zéninari, D. Courtois, G. Durry, A spectroscopic study of water vapor isotopologues $H_2{}^{16}O$, $H_2{}^{18}O$ and HDO using a continuous wave DFB quantum cascade laser in the 6.7 μm region for atmospheric applications, J. Quant. Spectrosc. Radiat. Transfer 102 (2006) 129–138.
94 F. Matsushima, N. Tomatsu, T. Nagai, Y. Moriwaki, K. Takagi, Frequency measurement of pure rotational transitions in the $v_2 = 1$ state of H_2O, J. Mol. Spectrosc. 235 (2006) 190–195.
95 G. Cazzoli, C. Puzzarini, G. Buffa, O. Tarrini, Pressure-broadening of water lines in the THz frequency region: Improvements and confirmations for spectroscopic databases. Part II, J. Quant. Spectrosc. Radiat. Transfer 110 (2009) 609–618.
96 B. J. Drouin, S. Yu, J. C. Pearson, H. Gupta, Terahertz spectroscopy for space applications: 2.5–2.7 THz spectra of HD, H_2O and NH_3, J. Mol. Struct. 1066 (2011) 2–12.
97 S. Yu, J. C. Pearson, B. J. Drouin, M.-A. Martin-Drumel, O. Pirali, M. Vervloet, L. H. Coudert, H. S. P. Müller, S. Brünken, Measurement and analysis of new terahertz and far-infrared spectra of high temperature water, J. Mol. Spectrosc. 279 (2012) 16–25.
98 Y. Lu, X.-F. Li, J. Wang, A.-W. Liu, S.-M. Hu, H_2O line positions in the 784–795 nm region with 10^{-9} accuracy, J. Quant. Spectrosc. Radiat. Transfer 118 (2013) 96–101.
99 L. Regalia, C. Oudot, S. Mikhailenko, L. Wang, X. Thomas, A. Jenouvrier, P. Von der Heyden, Water vapor line parameters from 6450 to 9400 cm^{-1}, J. Quant. Spectrosc. Radiat. Transfer 136 (2014) 119–136.
100 V. T. Sironneau, J. T. Hodges, Line shapes, positions and intensities of water transitions near 1.28 μm, J. Quant. Spectrosc. Radiat. Transfer 152 (2015) 1–15.
101 S. Kassi, T. Stoltmann, M. Casado, M. Daeron, A. Campargue, Lamb dip CRDS of highly saturated transitions of water near 1.4 μm, J. Chem. Phys 148 (2018) 054201.
102 S. N. Mikhailenko, D. Mondelain, E. V. Karlovets, S. Kassi, A. Campargue, Comb-Assisted Cavity Ring Down Spectroscopy of ^{17}O enriched water between 6667 and 7443 cm^{-1}, J. Quant. Spectrosc. Radiat. Transfer 206 (2018) 163–171.
103 J. K. G. Watson, Robust weighting in least-squares fits, J. Mol. Spectrosc. 219 (2003) 326–328.
104 W. Ritz, On a new law of series spectra, Astrophys. J. 28 (1908) 237.
105 J. K. G. Watson, The Use of Term-Value Fits in Testing Spectroscopic Assignments, J. Mol. Spectrosc. 165 (1994) 283 – 290.
106 G. Guennebaud, B. Jacob, et al., Eigen v3, http://eigen.tuxfamily.org, 2010.
107 T. Furtenbacher, J. Tennyson, O. V. Naumenko, O. L. Polyansky, N. F. Zobov, A. G. Császár, The 2018 Update of the IUPAC Database of Water Energy Levels, J. Quant. Spectrosc. Radiat. Transfer (In preparation).
108 R. A. Toth, $D_2{}^{16}O$ and $D_2{}^{18}O$ transition frequencies and strengths in the v_2 bands, J. Mol. Spectrosc. 162 (1993) 41–54.
109 J. Kauppinen, K. Jomena, V.-M. Horneman, New wavenumber calibration tables for H_2O, CO_2 and OCS lines between 400 cm^{-1} and 900 cm^{-1}, Appl. Opt. 21 (1982) 3332–3336.
110 C. Y. Lee, An algorithm for path connections and its applications, IRE Trans. Electron. Comput. (1961) 346–365.